Chapter 8

Regression II

Regression serves in this chapter to relate two climate variables, $X(i)$ and $Y(i)$. This is a standard tool for formulating a quantitative "climate theory" based on equations. Owing to the complexity of the climate system, such a theory can never be derived alone from the pure laws of physics—it requires to establish empirical relations between observed climate processes.

Since not only $Y(i)$ but also $X(i)$ are observed with error, the relation has to be formulated as an errors-in-variables model, and the estimation has to be carried out using adaptions of the OLS technique. This chapter focuses on the linear model and studies three estimation techniques (denoted as OLSBC, WLSXY and Wald–Bartlett procedure). It presents a novel bivariate resampling approach (pairwise-MBBres), which enhances the coverage performance of bootstrap CIs for the estimated regression parameters.

Monte Carlo simulations allow to assess the role of various aspects of the estimation. First, prior knowledge about the size of the measurement errors is indispensable to yield a consistent estimation. If this knowledge is not exact, which is typical for a situation in the climatological practice, it contributes to the estimation error of the slope (RMSE and CI length). This contribution persists even when the data size goes to infinity; the RMSE does then not approach zero. Second, autocorrelation has to be taken into account to prevent estimation errors unrealistically small and CIs too narrow.

This chapter studies two extensions of high relevance for climatological applications: linear prediction and lagged regression.

Regression as a method to estimate the trend in the climate equation (Eq. 1.2) is presented in Chapter 4.

M. Mudelsee, *Climate Time Series Analysis*, Atmospheric and Oceanographic Sciences Library 42, DOI 10.1007/978-90-481-9482-7_8,

8.1 Linear regression

To make a regression of the predictor variable, X, on the response variable, Y, we re-apply the errors-in-variables model (Section 4.1.7),

$$Y(i) = \beta_0 + \beta_1 \left[X(i) - S_X(i) \cdot X_{\text{noise}}(i)\right] + S_Y(i) \cdot Y_{\text{noise}}(i), \tag{8.1}$$

$i = 1, \ldots, n$. The variability of process $X(i)$ and $Y(i)$ is denoted as $S_X(i)$ and $S_Y(i)$, respectively; the noise component, $X_{\text{noise}}(i)$ and $Y_{\text{noise}}(i)$, is of assumed AR(1) type with persistence time τ_X and τ_Y, respectively. One task is to estimate the regression parameters, β_0 and β_1, given a bivariate sample, $\{t(i), x(i), y(i)\}_{i=1}^n$. Another, related task is to make a prediction of an unknown Y for a given value of X.

The errors-in-variables model (Eq. 8.1) differs from the simple model (Eq. 4.3) in its nonzero noise component of the predictor. Several estimators for the errors-in-variables model have been developed to deal with this more complex situation.

8.1.1 Ordinary least-squares estimation

The simple OLS estimation minimizes the unweighted sum of squares,

$$SSQ(\beta_0, \beta_1) = \sum_{i=1}^{n} \left[y(i) - \beta_0 - \beta_1 x(i)\right]^2 . \tag{8.2}$$

This yields the estimators

$$\widehat{\beta}_0 = \left[\sum_{i=1}^{n} y(i) - \widehat{\beta}_1 \sum_{i=1}^{n} x(i)\right] \Big/ n \tag{8.3}$$

and

$$\widehat{\beta}_1 = \left\{ \left[\sum_{i=1}^{n} x(i)\right] \left[\sum_{i=1}^{n} y(i)\right] \Big/ n - \sum_{i=1}^{n} x(i)\, y(i) \right\} \times \left\{ \left[\sum_{i=1}^{n} x(i)\right]^2 \Big/ n - \sum_{i=1}^{n} x(i)^2 \right\}^{-1} . \tag{8.4}$$

Using OLS means ignoring heteroscedasticity, persistence and errors in the predictor variable, X. However, heteroscedasticity and persistence can successfully be taken into account by employing WLS and GLS estimation, respectively. The success of ignoring errors in X depends on how large these are relative to the spread of the "true" X values (Eq.

4.34), which are given by $X_{\text{true}}(i) = X(i) - S_X(i) \cdot X_{\text{noise}}(i)$. If $S_X(i) = S_X$ is constant and $S_X^2 \ll VAR[X_{\text{true}}(i)]$, the estimation bias should be negligible. If $S_X(i)$ is not constant, one may expect a similar condition to the average of $S_X(i)$. The decisive quantity is $VAR[X_{\text{true}}(i)]$, which may be difficult to control for an experimenter prior to sampling the process.

If $X_{\text{noise}}(i)$ and $Y_{\text{noise}}(i)$ are independent, the estimator $\widehat{\beta}_1$ is biased downwards (Section 4.1.7) as $E\left(\widehat{\beta}_1\right) = \kappa \cdot \beta_1$, where $\kappa \leq 1$ is the attenuation factor or reliability ratio,

$$\kappa = \left(1 + S_X^2 \,/ VAR\left[X_{\text{true}}(i)\right]\right)^{-1}. \tag{8.5}$$

The intuitive reason of the bias downwards is that "smearing" the "true" predictor variable, $X_{\text{true}}(i)$, leads to a situation where the "cheapest fit solution" in terms of SSQ is a line that is horizontally tilted (Fig. 8.1).

8.1.1.1 Bias correction

Eq. (8.5) points to a bias-corrected slope estimation. Let $S_X(i) = S_X$ be constant and known, and let the variance of the "true" predictor values be given by $VAR[X_{\text{true}}(i)] = VAR[X(i)] - S_X^2$. This leads to

$$\widehat{\beta}_1 = \widehat{\beta}_{1,\text{OLS}} \,/ \left\{1 - S_X^2 \,/ VAR\left[X(i)\right]\right\}, \tag{8.6}$$

where $\widehat{\beta}_{1,\text{OLS}}$ is the simple OLS slope estimator (Eq. 8.4). We denote this estimation method (Eq. 8.6) as ordinary least squares with bias correction (OLSBC). The OLSBC intercept estimator equals the OLS intercept estimator (Eq. 8.3). In practice (sample level), plug in $x(i)$ for $X(i)$.

8.1.1.2 Prior knowledge about standard deviations

Assume homoscedastic noise components, $S_Y(i) = S_Y$ and $S_X(i) = S_X$, and denote their squared ratio as

$$\lambda = S_Y^2 \,/ S_X^2. \tag{8.7}$$

Knowledge prior to the estimation about S_X, S_Y or λ can increase the estimation accuracy.

If S_X is known, then OLSBC can be readily performed (Eq. 8.6). Such prior knowledge may be acquired, for example, by repeating measurements. Or there may exist theoretical information about the measuring device and, hence, S_X.

If S_X is only known within bounds, OLSBC estimation can still be applied. CI construction has then to take into account the limited prior

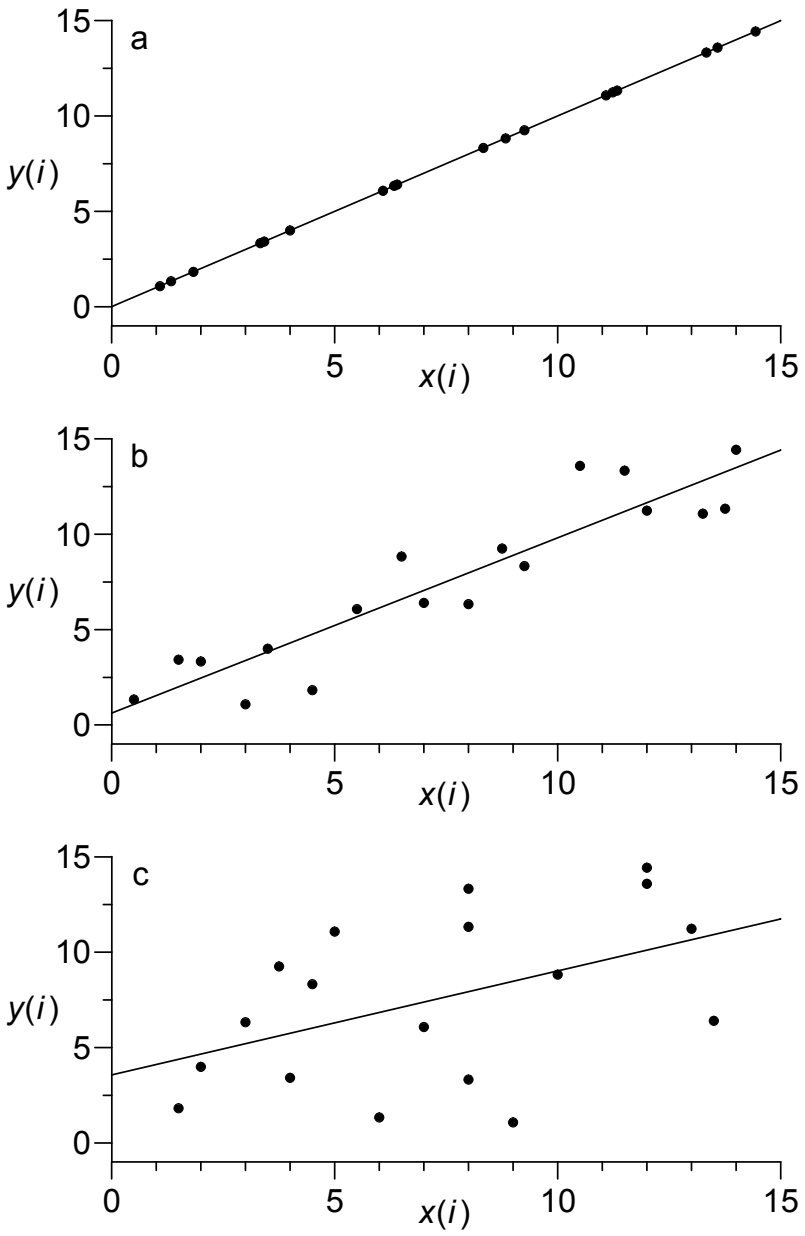

Figure 8.1. Linear errors-in-variables regression model, OLS estimation. The $\{y(i)\}_{i=1}^{n}$ are identical in panels **a–c;** the data size is $n = 18$; and the $\{x(i)\}_{i=1}^{n}$ are realizations of a predictor variable, $X(i)$, with constant zero (**a**), small (**b**) and large (**c**) noise components, $S_X(i) \cdot X_{\text{noise}}(i)$. The true slope is $\beta_1 = 1.0$ (**a**). The OLS fits (*solid lines*) exhibit slope estimates that are unbiased (**a** $\widehat{\beta}_1 = 1.0$) or biased (**b** $\widehat{\beta}_1 = 0.92$; **c** $\widehat{\beta}_1 = 0.55$).

knowledge. The result is a wider CI compared to the situation of perfect prior knowledge (Section 8.3).

If only the ratio, λ, is known, then one may be tempted to employ the method of moments estimator from the background material (Eq. 8.26) and plug in $\widehat{S}_X$ for S_X in Eq. (8.6). Similarly, if only S_Y is known, then one may be tempted to employ Eq. (8.26), replace therein $\delta = \lambda^{1/2}$ by $S_Y/\widehat{S}_X$ and solve the equation for $\widehat{S}_X$. However, own Monte Carlo experiments (results not shown) revealed completely inacceptable coverage accuracies of bootstrap confidence intervals for the slope (but acceptable accuracies for the intercept). The reason is the inaccurate $\widehat{S}_X$ estimation (Fuller 1987: Section 2.5 therein). Our recommendation for the case of known λ (or S_Y) is the weighted least-squares estimation (Section 8.1.2).

If no knowledge at all exists about S_X, S_Y or λ, then we face difficulties. One may simply try OLS but risk a biased slope estimation. One may resort to the Wald–Bartlett procedure (Section 8.1.3), but also this does not produce accurate results when so little is known. We discourage from adopting an OLS regression of Y an X and estimating S_X via the residual mean square (Eq. 4.8), an idea found occasionally in the literature. Own Monte Carlo experiments (similar to those in Section 8.3, results not shown) revealed inacceptable coverage performance of bootstrap CIs.

8.1.2 Weighted least-squares for both variables estimation

Studying the combination of both noise components in Eq. (8.1) in the form of $S_Y(i) \cdot Y_{\text{noise}}(i) - \beta_1 S_X(i) \cdot X_{\text{noise}}(i)$ makes clear the estimation approach via attaching weights to the observations of both variables (Deming 1943; Lindley 1947). The variant by York (1966) and others, who suggested minimization of the weighted least-squares sum,

$$SSQWXY(\beta_0, \beta_1) = \sum_{i=1}^{n} \frac{[y(i) - \beta_0 - \beta_1 x(i)]^2}{S_Y(i)^2 + \beta_1^2 S_X(i)^2}, \tag{8.8}$$

was included in the Numerical Recipes (Press et al. 1992: Section 15.3 therein). However, no general analytical solution exists and some numerical difficulties have to be circumnavigated (Section 8.8). We abbreviate this estimation procedure as WLSXY (Fig. 8.2).

8.1.2.1 Prior knowledge about standard deviation ratio

Assume $S_Y(i)$ and $S_X(i)$ to be unknown, but their (squared) ratio,

$$\lambda = S_Y(i)^2 \,/ S_X(i)^2, \tag{8.9}$$

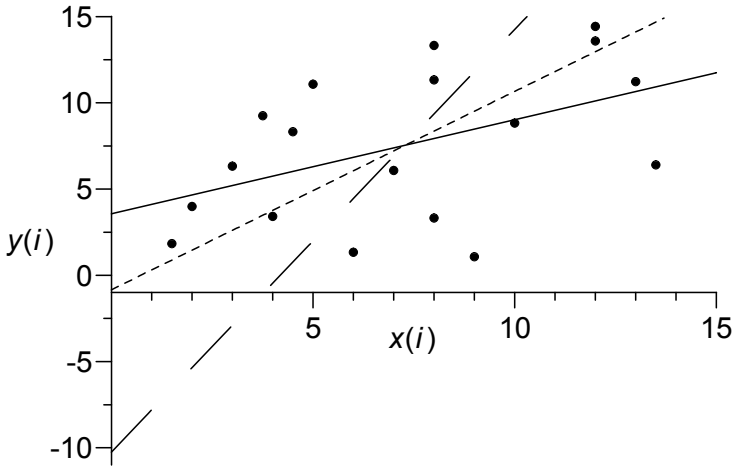

Figure 8.2. Linear errors-in-variables regression model, WLSXY and OLS estimations. The $\{x(i), y(i)\}_{i=1}^{n}$ are overtaken from Fig. 8.1c. The OLS fit of X on Y (*solid line*) has a slope of $\widehat{\beta}_1 = 0.55$, the OLS fit of Y on X (*long-dashed line*) has $1/\widehat{\beta}_1' = 2.45$ and the WLSXY fit of X on Y (*short-dashed line*) has $\widehat{\beta}_1 = 1.15$. (The model for the regression of Y on X is $X(i) = \beta_0' + \beta_1' Y(i) + S_X \cdot X_{\text{noise}}(i)$.)

to be constant and known. Such type of knowledge may be available in climatological applications. Then,

$$SSQWXY(\beta_0, \beta_1) = \sum_{i=1}^{n} \frac{[y(i) - \beta_0 - \beta_1 x(i)]^2}{\left(1 + \beta_1^2/\lambda\right) S_Y(i)^2}, \tag{8.10}$$

which is minimized (Section 8.8). The sub-case of constant $S_Y(i)$ (or $S_X(i)$) is considerably easier to treat than the heteroscedastic sub-case.

Under the assumption of Gaussian distributional shapes of $X_{\text{noise}}(i)$ and $Y_{\text{noise}}(i)$, the WLSXY estimators equal the maximum likelihood estimators (Madansky 1959; Fuller 1987).

8.1.2.2 Geometric interpretation

WLSXY minimizes the sum of squares of distances between the fit line and the data points. How to measure the distance depends on the ratio, $\lambda = S_Y(i)^2/S_X(i)^2$. The geometric interpretation is straightforward (Fig. 8.3) and generalizable to higher dimensions (background material).

If $S_X(i) = 0$, that means, the $X(i)$ values are exact, then $\lambda = \infty$ and we use WLS regression of X on Y (Section 4.1.1); if further $S_Y(i)$ is constant, this amounts to OLS regression. On the other hand, if $S_Y(i) = 0$, then $\lambda = 0$ and we use WLS regression of Y on X. (See Fig.

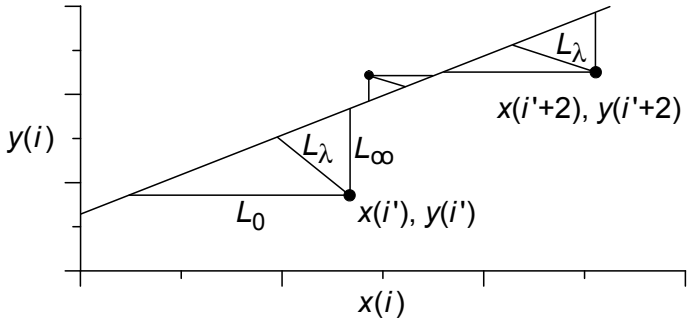

Figure 8.3. Geometric interpretation of WLSXY. The lines L_0, L_λ and L_∞ measure the distance from a data point to the fit line for $\lambda = 0$, $0 < \lambda < \infty$ and $\lambda = \infty$, respectively.

8.2 for the regression of Y on X.) If the standard deviations are nonzero and $0 < \lambda < \infty$, we measure the distance along the line L_λ (Fig. 8.3). The slope of this line is equal to $-\lambda/\widehat{\beta}_1$ (York 1967).

If heteroscedasticity is in one or both of the noise components, then the ratio λ may vary with time (i) and, hence, the line L_λ may vary in its slope. The difficulty of non-identifiability is introduced by unknown λ because then it is not unequivocally determined how to measure the distance and minimize the sum of squares.

8.1.3 Wald–Bartlett procedure

A straightforward estimation idea (Draper and Smith 1981: Section 2.14 therein) is to build two groups of the bivariate sample according to the size of the x values, then to take for each group the centres defined by the x and y averages and, finally, to connect the centres using a straight line—defining the estimate of the slope. The intercept estimate is found via the centre of the complete bivariate sample and the slope estimate. This goes back to Wald (1940), who grouped the sample into two halves of same size (if n is even) and Bartlett (1949), who showed that taking three groups improves the accuracy of the regression estimators. (Intuitively, the means of the two groups are further apart for taking thirds than for taking halves, outweighing the deficit of reduced data sizes.) We call this estimation Wald–Bartlett procedure (Fig. 8.4).

The Wald–Bartlett procedure can in principle be applied to *any* grouping of the set of data points, not only according to the size of the x values. A point to note is that the grouping has to be independent of $X_{\text{noise}}(i)$ for achieving consistency of the estimators (Wald 1940). This condition is violated when the $\{X_{\text{true}}(i)\}_{i=1}^{n}$ are unknown and the size ordering

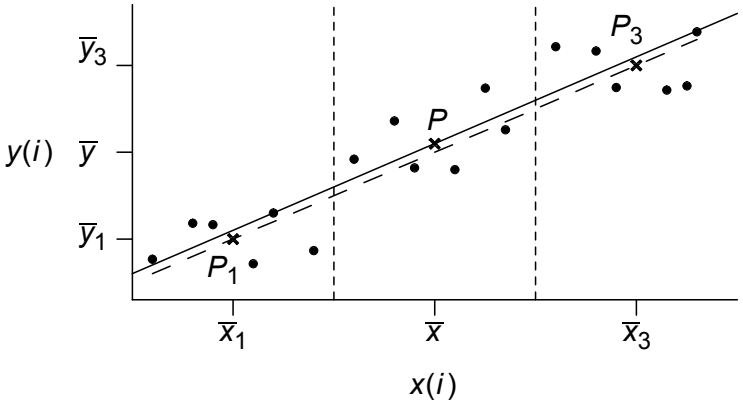

Figure 8.4. Wald–Bartlett procedure. The bivariate sample $\{x(i), y(i)\}_{i=1}^{n}$ is divided into three groups of same size according to the size of the x values; if n is not divisible by 3, then take the closest grouping. Let j index the size-sorted sample. Let the averages of $\{x(j)\}_{j=1}^{n/3}$ and $\{y(j)\}_{j=1}^{n/3}$, denoted as $\bar{x}_1$ and $\bar{y}_1$, define the first group's centre (P_1, *cross*), and let the averages of $\{x(j)\}_{j=2n/3+1}^{n}$ and $\{y(j)\}_{j=2n/3+1}^{n}$, denoted as $\bar{x}_3$ and $\bar{y}_3$, define the third group's centre (P_3). The line $P_1 P_3$ (*long-dashed*) defines the Wald–Bartlett regression estimate of the slope, $\widehat{\beta}_1 = (\bar{y}_3 - \bar{y}_1)/(\bar{x}_3 - \bar{x}_1)$. The centre of the complete sample (P) is defined via the averages of $\{x(j)\}_{j=1}^{n}$ and $\{y(j)\}_{j=1}^{n}$, denoted as $\bar{x}$ and $\bar{y}$. The Wald–Bartlett intercept estimate, $\widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1 \bar{x}$, completes the linear fit (*solid line*).

is made on the noise-influenced observations. Monte Carlo simulations, similar to those in Section 8.3, reveal that the inconsistency leads to an inacceptably poor coverage performance of bootstrap CIs (for $\widehat{\beta}_0$ and $\widehat{\beta}_1$) (not shown). This limits severely the applicability of the Wald–Bartlett procedure to real-world climatological problems, where the $X_{\text{true}}(i)$ are usually unknown.

Wald (1940: p. 298 therein) notes that if prior knowledge exists on the standard deviation ratio, then a consistent estimation could be constructed. This situation is similar to WLSXY estimation (Section 8.1.2.1).

The calculation of classical CIs (Wald 1940; Bartlett 1949) via the Student's t distribution assumed prior knowledge to be available, allowing a consistent estimation, and the errors, $X_{\text{noise}}(i)$ and $Y_{\text{noise}}(i)$, to be serially independent and of Gaussian shape.

8.2 Bootstrap confidence intervals

Classical CIs are based on the PDF of an estimator (Chapter 3). The PDF can be analytically determined unless the situation (estimation

problem, noise properties) becomes too complex. The construction of classical CIs for the linear errors-in-variables model (Wald 1940; Bartlett 1949; York 1966; Fuller 1987) made a number of assumptions from the following:

1. Gaussian distributional shapes of the noise components, $X_{\text{noise}}(i)$ and $Y_{\text{noise}}(i)$;
2. absence of autocorrelation in the noise components;
3. absence of correlation between $X(i)$ and $X_{\text{noise}}(i)$ as well as between $Y(i)$ and $Y_{\text{noise}}(i)$;
4. absence of correlation between $X_{\text{noise}}(i)$ and $Y_{\text{noise}}(i)$.

Some authors treat the correlation effects (points 3 and 4) and non-Gaussian errors (point 1), see the background material (Section 8.7). However, allowance for autocorrelations (point 2) seems to have been made by none.

Here we are interested in linearly relating two climate processes, $X(i)$ and $Y(i)$, and our sample, $\{t(i), x(i), y(i)\}_{i=1}^{n}$, contains the time. The previous chapters document that non-Gaussian distributions and persistence phenomena are typical of climate processes. We cannot therefore expect the classical method to yield accurate results for climate data. This is, as in previous chapters, the reason to consider the bootstrap method. An additional point is incomplete knowledge about the noise components. Often we have no or only limited information about $S_Y(i)$, $S_X(i)$ or their (squared) ratio, λ. Such incomplete knowledge, which may widen the CI, is quantifiable using bootstrap resampling (Booth and Hall 1993).

One resampling algorithm is the pairwise-MBB, which has been found useful in the context of correlation estimation (Algorithm 7.2).

The other algorithm, introduced here for the purpose of enhancing the coverage performance in the context of fitting errors-in-variables models, is called pairwise-moving block bootstrap resampling of residuals or pairwise-MBBres. It is based on the observation that the (linear) errors-in-variables regression (Eq. 8.1) is a model with a deterministic (linear) component. Since pairwise resampling seems to be handicapped in the presence of deterministic components (Chapter 4), the idea of the pairwise-MBBres algorithm is to take the fit and the regression residuals, apply pairwise-MBB to the residuals and add the resampled residuals to the fit. The new approach is that the residuals, $e_X(i)$ and $e_Y(i)$, are in two dimensions (Fig. 8.5). The pairwise-MBBres is given as Algorithm 8.1.

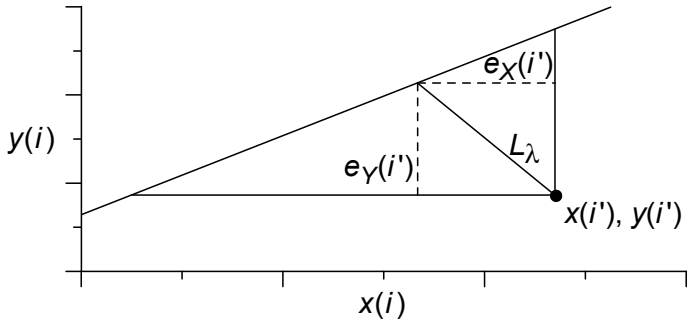

Figure 8.5. Pairwise-MBBres algorithm, definition of residuals. The line L_λ measures the distance from a data point to the fit line (Fig. 8.3). The residuals (*dashed lines*) are given by $e_X(i) = [\widehat{\beta}_0 + \widehat{\beta}_1 \cdot x(i) - y(i)]/[\lambda/\widehat{\beta}_1 + \widehat{\beta}_1]$ and $e_Y(i) = -\lambda \cdot e_X(i)/\widehat{\beta}_1$.

8.2.1 Simulating incomplete prior knowledge

Assume for the convenience of exposition homoscedastic noise components, $S_X(i) = S_X$ and $S_Y(i) = S_Y$. For achieving an identifiable problem, OLSBC estimation requires information, not contained in the sample, about S_X (Section 8.1.1.2); analogously, WLSXY estimation requires information about both S_X and S_Y, or about their ratio, $\delta = \lambda^{1/2} = S_Y/S_X$ (Section 8.1.2.1).

In practical applications such prior knowledge is not always exact. S_X or δ are then described by random variables. Bootstrap resampling can be augmented by a simulation step, where random numbers are drawn from the distribution of S_X or δ. This increases the uncertainty of the OLSBC or WLSXY estimates, leading to wider bootstrap CIs compared to a situation with exact prior knowledge (Booth and Hall 1993).

In the Monte Carlo experiments studying incomplete prior knowledge, we use the model

$$\sqrt{\lambda^*} = \delta^* = \delta \cdot \mathcal{E}_{\mathrm{U}[1-\Delta;\,1+\Delta]}(i), \tag{8.11}$$

$$S_X^* = S_X/\sqrt{\delta^*}, \tag{8.12}$$

where $\mathcal{E}_{\mathrm{U}[1-\Delta;\,1+\Delta]}(i)$ is an IID random process with a uniform distribution over the interval $[1-\Delta;\ 1+\Delta]$. For example, $\Delta = 0.5$ specifies a situation where we only know δ to lie between 0.5 and 1.5 times its true value. Other models are possible.

The construction of bootstrap CIs (Algorithm 8.1) is adapted at Step 8, the calculation of the replications. Instead of applying to the resampled data the same estimation procedure that is used for the original data, an adapted estimation is performed (Steps 8a and 8b).

Step 1	Bivariate time series	$\{t(i), x(i), y(i)\}_{i=1}^{n}$
Step 2	Parameter estimates	$\widehat{\beta}_0, \widehat{\beta}_1$
	from OLSBC, WLSXY or	
	Wald–Bartlett procedure	
Step 3	Residuals (Fig. 8.5)	$e_X(i), e_Y(i)$
Step 4	Fit values	$x_{\text{fit}}(i) = x(i) - e_X(i),$
		$y_{\text{fit}}(i) = y(i) - e_Y(i)$
Step 5	Bias-corrected AR(1)	
	parameters,	$\widehat{a}'_X = \widehat{a}'_Y$
	estimated on residuals,	
	block length selection	l
	after Eqs. (7.31) and (7.32)	
Step 6	Resampled residuals,	
	pairwise-MBB with l	$\left\{e_X^{*b}(i), e_Y^{*b}(i)\right\}_{i=1}^{n}$ (b, counter)
Step 7	Resample	$x^{*b}(i) = x_{\text{fit}}(i) + e_X^{*b}(i),$
		$y^{*b}(i) = y_{\text{fit}}(i) + e_Y^{*b}(i), i = 1, \ldots, n$
Step 8	Bootstrap replications	$\widehat{\beta}_0^{*b}, \widehat{\beta}_1^{*b}$
Step 9	Bootstrap prediction	$\widehat{y}^{*b}(n+1) = \widehat{\beta}_0^{*b} + \widehat{\beta}_1^{*b} x(n+1)$
Step 10	Go to Step 6 until $b = B$	
	(usually $B = 2000$)	
	replications exist	$\left\{\widehat{\beta}_0^{*b}\right\}_{b=1}^{B}, \left\{\widehat{\beta}_1^{*b}\right\}_{b=1}^{B}, \left\{\widehat{y}^{*b}(n+1)\right\}_{b=1}^{B}$
Step 11	Calculate CIs	
	(Section 3.4)	

Algorithm 8.1. Construction of bootstrap confidence intervals for parameters of the linear errors-in-variables regression model, pairwise-MBBres resampling, even spacing. In case of uneven spacing, Step 5 uses $\widehat{\tau}'_X = \widehat{\tau}'_Y$. Step 8 can be adapted as follows for taking incomplete prior knowledge into account. Step 8a: simulate λ^*, S_X^*; Step 8b: use OLSBC with S_X^* instead of S_X (Eq. 8.6), use WLSXY with λ^* instead of λ (Eq. 8.10). Steps 9 and 10: $\widehat{y}^{*b}(n+1)$ refers to prediction (Section 8.5).

8.3 Monte Carlo experiments

The first group of experiments (Section 8.3.1) adopts an "easy" setting, where distributional shapes are Gaussian and prior knowledge is exact. The results confirm the success of block resampling methods in preserving serial dependence. The second group of experiments (Section 8.3.2) shows that for realistic settings, with non-Gaussian distributions or incomplete knowledge, the results are less exact. It appears that then WLSXY estimation combined with pairwise-MBBres resampling yields the relatively best results. The third group (Section 8.3.3) quantifies coverage accuracy and RMSE in dependence on the accuracy of the prior knowledge (standard deviation ratio). It demonstrates that even with $n \to \infty$, the RMSE (and the CI length) for $\widehat{\beta}_1$ (slope) does not go to zero, but rather approaches finite values. On the other hand (intercept), with $n \to \infty$ does $\mathrm{RMSE}_{\widehat{\beta}_0} \to 0$. The last group (Section 8.3.4), finally, explores what happens when we mis-specify the degree of how accurately we know the standard deviation ratio.

8.3.1 Easy setting

The easy setting (Gaussian shapes of $X_{\mathrm{noise}}(i)$ and $Y_{\mathrm{noise}}(i)$, complete prior knowledge) is further simplified when no autocorrelation resides in the noise components. Table 8.1 exhibits excellent coverage performance of bootstrap CIs with pairwise-MBBres resampling already for sample sizes as small as 20. The excellent performance regards both parameters (intercept and slope) and all estimation procedures (OLSBC, WLSXY and Wald–Bartlett). Similar results for OLSBC and WLSXY were obtained using pairwise-MBB resampling (results not shown).

If there exists autocorrelation, then pairwise-MBBres resampling successfully preserves it, where it does not matter whether the AR(1) parameters are known or, which is more realistic, have to be estimated. However, there is one exception: the Wald–Bartlett estimation of the intercept fails completely, independent of the sample size. Also OLSBC and WLSXY estimations of the intercept are of reduced CI coverage accuracy, but become acceptable for n above, say, 200. It is not clear why intercept estimation is more problematic than slope estimation.

We remark that the inacceptable performance of the Wald–Bartlett procedure occurred even in the presence of knowledge of the size of the true predictor values, which in turn enabled a perfect grouping (independent of the predictor noise). The three points of the (1) inacceptable performance of CIs for the intercept, (2) not better (compared with OLSBC and WLSXY) performance of CIs for the slope and (3) rather strong requirement of knowledge of the size of true predictor values—

Table 8.1. Monte Carlo experiment, linear errors-in-variables regression with AR(1) noise of normal shape and complete prior knowledge: CI coverage performance. $n_{\text{sim}} = 47{,}500$ random samples were generated from $X_{\text{true}}(i) = \mathcal{E}_{\text{N}(0,\,1)}(i)$ and $Y(i) = \beta_0 + \beta_1 X_{\text{true}}(i) + S_Y \cdot Y_{\text{noise}}(i)$, $i = 1, \ldots, n$, with $\beta_0 = 1.0$ and $\beta_1 = 2.0$. Predictor noise was subsequently added, $X(i) = X_{\text{true}}(i) + S_X \cdot X_{\text{noise}}(i)$. The $X_{\text{noise}}(i)$ and $Y_{\text{noise}}(i)$ are mutually independent Gaussian AR(1) processes for even spacing (Eq. 2.1) with parameters a_X and a_Y, respectively. Construction of bootstrap CIs used pairwise-MBBres resampling (Algorithm 8.1), block length selection after Eqs. (7.31) and (7.32), the Student's t interval type ($\nu = n - 2$), $B = 2000$ and confidence level 95%. Prior knowledge of $S_X = 0.25, S_Y = 0.5$ and the size of the $\{x_{\text{true}}(i)\}_{i=1}^{n}$ was exact and utilized in the estimations; AR(1) parameters are in two cases known, in one case unknown and estimated with bias correction.

n	$\gamma_{\widehat{\beta}_0}$[a]			$\gamma_{\widehat{\beta}_1}$[a]			*Nominal*
	Estimation method			*Estimation method*			
	OLSBC	*WLSXY*	*WB*[b]	*OLSBC*	*WLSXY*	*WB*[b]	
	$a_X = a_Y = 0.0$ (*known*)						
10	0.933	0.928	0.933	0.955	0.928	0.938	0.950
20	0.939	0.939	0.942	0.949	0.938	0.942	0.950
50	0.944	0.947	0.948	0.949	0.946	0.947	0.950
100	0.944	0.946	0.947	0.949	0.947	0.946	0.950
200	0.945	0.949	0.950	0.949	0.947	0.945	0.950
500	0.946	0.949	0.949	0.950	0.946	0.946	0.950
1000	0.945	0.948	0.949	0.949	0.948	0.945	0.950
	$a_X = a_Y = 0.3$ (*known*)						
10	0.839	0.831	0.785	0.949	0.919	0.912	0.950
20	0.863	0.862	0.816	0.947	0.936	0.926	0.950
50	0.895	0.896	0.827	0.948	0.945	0.930	0.950
100	0.909	0.911	0.827	0.947	0.945	0.933	0.950
200	0.921	0.924	0.836	0.947	0.945	0.939	0.950
500	0.931	0.933	0.840	0.948	0.944	0.941	0.950
1000	0.935	0.938	0.844	0.950	0.949	0.945	0.950
	$a_X = a_Y = 0.3$ (*unknown*, *estimated*)						
10	0.807	0.797	0.841	0.922	0.892	0.935	0.950
20	0.871	0.869	0.853	0.943	0.932	0.947	0.950
50	0.896	0.897	0.852	0.946	0.943	0.947	0.950
100	0.911	0.913	0.854	0.948	0.947	0.949	0.950
200	0.921	0.923	0.853	0.949	0.948	0.949	0.950
500	0.930	0.932	0.850	0.949	0.945	0.948	0.950
1000	0.934	0.937	0.849	0.949	0.948	0.949	0.950

[a] Standard errors of $\gamma_{\widehat{\beta}_0}$ and $\gamma_{\widehat{\beta}_1}$ are nominally 0.001.

[b] Wald–Bartlett procedure.

Table 8.2. Monte Carlo experiment, linear errors-in-variables regression with AR(1) noise of normal shape and complete prior knowledge: CI coverage performance (continued). Design is identical to the previous experiment (Table 8.1), with the exception that autocorrelation is ignored at CI construction.

n	$\gamma^{\mathrm{a}}_{\widehat{\beta}_0}$			$\gamma^{\mathrm{a}}_{\widehat{\beta}_1}$			*Nominal*
	Estimation method			*Estimation method*			
	OLSBC	*WLSXY*	*WB*[b]	*OLSBC*	*WLSXY*	*WB*[b]	
	$a_X = a_Y = 0.3$ (*ignored*)						
10	0.846	0.838	0.846	0.955	0.928	0.939	0.950
20	0.842	0.840	0.845	0.948	0.937	0.942	0.950
50	0.845	0.847	0.849	0.948	0.944	0.946	0.950
100	0.842	0.846	0.846	0.947	0.946	0.946	0.950
200	0.845	0.848	0.849	0.947	0.946	0.945	0.950
500	0.846	0.849	0.849	0.948	0.945	0.947	0.950
1000	0.846	0.848	0.849	0.946	0.948	0.946	0.950

[a] Standard errors of $\gamma_{\widehat{\beta}_0}$ and $\gamma_{\widehat{\beta}_1}$ are nominally 0.001.
[b] Wald–Bartlett procedure.

provide enough support to exclude the Wald–Bartlett procedure from consideration in the further experiments (which have more realistic settings).

One experiment under the easy setting studied what happens when autocorrelation is ignored (Table 8.2). This was achieved by prescribing positive AR(1) parameters, a_X and a_Y, of both noise components and resetting their estimate values to zero, $\widehat{a}'_X \equiv 0$ and $\widehat{a}'_Y \equiv 0$. The detrimental effect, again on $\widehat{\beta}_0$ but not $\widehat{\beta}_1$, was an underestimated bootstrap standard error, which led to too narrow CIs and too low coverages. (Similar results where found when replacing pairwise-MBBres by pairwise-MBB resampling.)

The major findings of the first experiment on CI coverage accuracy (Table 8.1) are reflected in the results on empirical RMSE (Table 8.3). Autocorrelation increases the estimation error of the intercept, but not of the slope. A larger data size means a smaller estimation error of the intercept and the slope. For $n \to \infty$, both $\mathrm{RMSE}_{\widehat{\beta}_0}$ and $\mathrm{RMSE}_{\widehat{\beta}_1}$ appear to go to zero.

OLSBC and WLSXY estimation methods perform similarly; only for small n and slope estimation does WLSXY seem to give smaller error bars than OLSBC.

Table 8.3. Monte Carlo experiment, linear errors-in-variables regression with AR(1) noise of normal shape and complete prior knowledge: RMSE. Design is identical to the experiment shown in Table 8.1.

n	$\text{RMSE}^{\text{a}}_{\widehat{\beta}_0}$		$\text{RMSE}^{\text{b}}_{\widehat{\beta}_1}$	
	Estimation method		*Estimation method*	
	OLSBC	*WLSXY*	*OLSBC*	*WLSXY*
	$a_X = a_Y = 0.0$			
10	0.237	0.245	0.645	0.289
20	0.161	0.164	0.190	0.178
50	0.099	0.101	0.110	0.105
100	0.070	0.071	0.076	0.073
200	0.049	0.050	0.053	0.052
500	0.031	0.032	0.033	0.033
1000	0.022	0.022	0.023	0.023
	$a_X = a_Y = 0.3$			
10	0.300	0.310	5.759	0.275
20	0.211	0.216	0.188	0.174
50	0.134	0.137	0.110	0.105
100	0.094	0.096	0.076	0.073
200	0.067	0.068	0.053	0.052
500	0.042	0.043	0.034	0.033
1000	0.030	0.031	0.024	0.023

[a] Empirical $\text{RMSE}_{\widehat{\beta}_0}$, given by $\left[\sum_{i=1}^{n_{\text{sim}}} \left(\widehat{\beta}_0 - \beta_0\right)^2 / n_{\text{sim}}\right]^{1/2}$.

[b] Empirical $\text{RMSE}_{\widehat{\beta}_1}$, given by $\left[\sum_{i=1}^{n_{\text{sim}}} \left(\widehat{\beta}_1 - \beta_1\right)^2 / n_{\text{sim}}\right]^{1/2}$.

8.3.2 Realistic setting: incomplete prior knowledge

The setting becomes more complex, or realistic, when the prior knowledge about the standard deviations of the measurement noise is not complete. We study (Table 8.4) a situation where the true ratio is $\delta = S_Y/S_X = 2.0$ but one knows only that δ is between 1.0 and 3.0. The adapted bootstrap CI construction (WLSXY with $\lambda^* = (\delta^*)^2$), for both $\widehat{\beta}_0$ and $\widehat{\beta}_1$, yields acceptable accuracies for normal shape and $n \gtrapprox 200$—under the condition that pairwise-MBBres resampling is employed. (For climatological purposes, a 95% CI may be "acceptable" if the true coverage is between, say, 92 and 98%.) The pairwise-MBBres resampling method (Fig. 8.5) is clearly superior to the pairwise-MBB method.

Bootstrap CI construction for OLSBC estimates failed to achieve the accuracies for WLSXY—for both resampling methods. Rather large

Table 8.4. Monte Carlo experiment, linear errors-in-variables regression with AR(1) noise of normal/lognormal shape and incomplete prior knowledge: CI coverage performance. Design is identical to the first experiment (Table 8.1), with the following exceptions: (1) autocorrelation parameters are unknown (and estimated with bias correction) and (2) $Y_{\text{noise}}(i)$ has normal or lognormal shape. Estimation and CI construction is identical to the first experiment (Table 8.1), with the following exceptions: (1) the Wald–Bartlett procedure is omitted; (2) prior knowledge of $S_X = 0.25, S_Y = 0.5$ ($\delta = 2.0$) is incomplete after Eqs. (8.11) and (8.12) with $\Delta = 0.5$; and (3) CI construction is adapted accordingly (Section 8.2.1).

n	$\gamma^{\text{a}}_{\hat{\beta}_0}$		$\gamma^{\text{a}}_{\hat{\beta}_1}$		*Nominal*
	Estimation method		*Estimation method*		
	OLSBC	*WLSXY*	*OLSBC*	*WLSXY*	
	$a_X = a_Y = 0.3$, $Y_{\text{noise}}(i)$ *normal shape*, *pairwise-MBBres*				
10	0.809	0.802	0.941	0.895	0.950
20	0.871	0.871	0.956	0.931	0.950
50	0.898	0.899	0.955	0.940	0.950
100	0.909	0.913	0.952	0.942	0.950
200	0.921	0.924	0.948	0.943	0.950
500	0.930	0.934	0.939	0.947	0.950
1000	0.937	0.939	0.927	0.952	0.950
2000	0.936	0.939	0.919	0.958	0.950
5000	0.941	0.944	0.909	0.960	0.950
	$a_X = a_Y = 0.3$, $Y_{\text{noise}}(i)$ *normal shape*, *pairwise-MBB*				
10	0.856	0.875	0.980	0.947	0.950
20	0.864	0.867	0.972	0.944	0.950
50	0.862	0.862	0.958	0.943	0.950
100	0.866	0.866	0.955	0.943	0.950
200	0.865	0.865	0.954	0.948	0.950
500	0.871	0.871	0.943	0.947	0.950
1000	0.873	0.871	0.932	0.952	0.950
2000	0.880	0.880	0.922	0.957	0.950
5000	0.891	0.889	0.915	0.962	0.950
	$a_X = 0.3, a_Y = 0.8$, $Y_{\text{noise}}(i)$ *lognormal shape*, *pairwise-MBBres*				
10	0.689	0.673	0.942	0.871	0.950
20	0.738	0.738	0.956	0.902	0.950
50	0.788	0.793	0.961	0.904	0.950
100	0.817	0.824	0.960	0.897	0.950
200	0.849	0.856	0.959	0.897	0.950
500	0.880	0.887	0.949	0.902	0.950
1000	0.894	0.901	0.936	0.917	0.950
2000	0.912	0.919	0.925	0.929	0.950
5000	0.924	0.931	0.917	0.947	0.950

[a] Standard errors of $\gamma_{\hat{\beta}_0}$ and $\gamma_{\hat{\beta}_1}$ are nominally 0.001.

sample sizes ($n = 2000$ and 5000) reveal the "worrisome" behaviour of $\gamma_{\widehat{\beta}_1}$ for the OLSBC estimates: they do not saturate and approach the nominal value of 0.95 but seem rather to drift away for large n.

It becomes clear that for realistic settings (autocorrelation, incomplete prior knowledge), WLSXY estimation combined with pairwise-MBBres resampling is the only one-loop option to achieve acceptable levels of CI accuracy. A second loop of resampling (calibration or bootstrap-t) may in principle improve the accuracy, also for errors-in-variables regression (Booth and Hall 1993).

The combination of WLSXY and pairwise-MBBres performed well (Table 8.4) also for a rather difficult setting (stronger, unequal autocorrelations, lognormal shape). It is interesting to note that slope estimation yielded more accurate results than intercept estimation. The data size requirements, however, become rather strong (Table 8.4). Obtaining accurate results for data sizes in the range of 500 and below may require calibration methods.

8.3.3 Dependence on accuracy of prior knowledge

In practical situations, our prior knowledge about the measurement standard errors or their ratio, $\delta = S_Y/S_X$, may depend to a considerable degree on how good we know the measurement devices (calibration standards, replication analyses, etc.) or the archives "containing" the data (sampling error). The accuracy of that knowledge, parameterized here in form of Δ (Eqs. 8.11 and 8.12), should influence the estimation RMSE and possibly also the CI accuracy. This is explored by means of a set of simulation experiments (Tables 8.5 and 8.6), where Δ is varied.

The selection of the other setting parameters follows the previous Monte Carlo experiments in this section: intermediate sizes of autocorrelation, Gaussian shape and a true standard deviation ratio of $\delta = 2.0$. The data size may take relatively large values ($n = 2000$ and 5000) because also the limiting behaviour is of interest. We employ the WLSXY estimation and Student's t CI constructed by means of pairwise-MBBres resampling.

The resulting coverages (Table 8.5) approach with increasing data size the nominal value—as they should. In general, the levels are acceptable from n above, say, 200 (slope estimation) or 500 (intercept estimation). In the case of slope estimation, a highly inaccurate prior knowledge ($\Delta = 0.9$) may require more data points for achieving a coverage level similar to values found for smaller inaccuracies ($\Delta \leq 0.7$).

The resulting RMSE values (Table 8.6) for intercept estimation approach zero with increasing data size. The rate of this convergence seems not to depend on the accuracy of the prior knowledge (Δ). The RMSE

Table 8.5. Monte Carlo experiment, linear errors-in-variables regression with AR(1) noise of normal shape: influence of accuracy of prior knowledge on CI coverage performance. Design is identical to the previous experiment (Table 8.4), with the following fixed setting: (1) autocorrelation parameters are $a_X = a_Y = 0.3$, (2) both noise components have normal shape. Estimation and CI construction is identical to the previous experiment (Table 8.4), with the following exceptions: (1) only WLSXY estimation with pairwise-MBBres resampling is considered; (2) prior knowledge of $S_X = 0.25, S_Y = 0.5$ ($\delta = 2.0$) is incomplete after Eq. (8.11) with various Δ values.

n	γ[a]					*Nominal*
	Accuracy of prior knowledge					
	$\Delta = 0.1$	$\Delta = 0.3$	$\Delta = 0.5$	$\Delta = 0.7$	$\Delta = 0.9$	
	Intercept estimation					
10	0.774	0.772	0.802	0.773	0.782	0.950
20	0.868	0.867	0.871	0.869	0.873	0.950
50	0.895	0.898	0.899	0.897	0.902	0.950
100	0.913	0.912	0.913	0.914	0.915	0.950
200	0.924	0.924	0.924	0.925	0.928	0.950
500	0.931	0.933	0.934	0.934	0.934	0.950
1000	0.936	0.936	0.939	0.939	0.941	0.950
2000	0.940	0.939	0.939	0.942	0.944	0.950
5000	0.945	0.945	0.944	0.944	0.946	0.950
	Slope estimation					
10	0.873	0.877	0.895	0.877	0.875	0.950
20	0.933	0.931	0.931	0.925	0.916	0.950
50	0.944	0.942	0.940	0.936	0.916	0.950
100	0.948	0.945	0.942	0.936	0.918	0.950
200	0.947	0.948	0.943	0.938	0.920	0.950
500	0.947	0.948	0.947	0.942	0.924	0.950
1000	0.948	0.947	0.952	0.945	0.925	0.950
2000	0.946	0.950	0.958	0.946	0.924	0.950
5000	0.950	0.963	0.960	0.945	0.923	0.950

[a] Standard errors of γ for intercept and slope estimations are nominally 0.001.

values for slope estimation show an interesting behaviour: they do not vanish with increasing data size but rather approach a finite value. The reason is that the inaccurate prior knowledge about the measurement standard errors (nonzero Δ) persists to influence the slope estimation—an error source independent of the data size. Similar behaviours were found also for OLSBC estimation of intercept and slope (results not shown). The saturation value of $\mathrm{RMSE}_{\widehat{\beta}_1}$ depends on the accuracy of the prior knowledge (Δ), seemingly in a close-to-linear relation.

Table 8.6. Monte Carlo experiment, linear errors-in-variables regression with AR(1) noise of normal shape: influence of accuracy of prior knowledge on RMSE. The experiment is the same as described in Table 8.5.

n	RMSE				
	Accuracy of prior knowledge				
	$\Delta = 0.1$	$\Delta = 0.3$	$\Delta = 0.5$	$\Delta = 0.7$	$\Delta = 0.9$
	Intercept estimation[a]				
10	0.310	0.313	0.315	0.329	0.348
20	0.217	0.219	0.220	0.227	0.241
50	0.137	0.137	0.139	0.143	0.151
100	0.096	0.097	0.098	0.100	0.106
200	0.068	0.069	0.069	0.071	0.075
500	0.043	0.043	0.044	0.045	0.048
1000	0.031	0.031	0.031	0.032	0.034
2000	0.022	0.022	0.022	0.023	0.024
5000	0.014	0.014	0.014	0.014	0.015
	Slope estimation[b]				
10	0.279	0.300	0.428	0.303	0.339
20	0.173	0.177	0.181	0.191	0.223
50	0.105	0.107	0.114	0.126	0.166
100	0.074	0.077	0.085	0.099	0.146
200	0.052	0.056	0.066	0.084	0.135
500	0.033	0.040	0.052	0.073	0.127
1000	0.024	0.032	0.047	0.069	0.125
2000	0.018	0.028	0.043	0.067	0.123
5000	0.012	0.025	0.042	0.066	0.124

[a] Empirical $\mathrm{RMSE}_{\widehat{\beta}_0}$, given by $\left[\sum_{i=1}^{n_{\mathrm{sim}}} \left(\widehat{\beta}_0 - \beta_0\right)^2 / n_{\mathrm{sim}}\right]^{1/2}$.

[b] Empirical $\mathrm{RMSE}_{\widehat{\beta}_1}$, given by $\left[\sum_{i=1}^{n_{\mathrm{sim}}} \left(\widehat{\beta}_1 - \beta_1\right)^2 / n_{\mathrm{sim}}\right]^{1/2}$.

To summarize, measurement error in the predictor requires to modify the OLS method to yield a bias-free slope estimation: OLSBC or WLSXY. These modified estimation methods require prior knowledge about the size of the measurement error. If this knowledge is not exact, which is a typical situation in the climatological practice, then it contributes to the estimation error of the slope (RMSE and CI length). This contribution persists even when the data size goes to infinity.

8.3.4 Mis-specified prior knowledge

What happens if we make a wrong specification of the accuracy of our prior knowledge? We study (Table 8.7) a situation where (1) the

Table 8.7. Monte Carlo experiment, linear errors-in-variables regression with AR(1) noise of normal shape: influence of mis-specified prior knowledge on CI coverage performance. Design and estimation (WLSXY) are identical to the previous experiment (Table 8.5). CI construction (via pairwise-MBBres resampling) is identical to that in the previous experiment, with the following exceptions: (1) prior knowledge of $S_X = 0.25, S_Y = 0.5$ ($\delta = 2.0$) is incomplete after Eq. (8.11) with $\Delta = 0.5$; (2) the adaptive Steps 8a and 8b of Algorithm 8.1 are allowed to mis-specify Δ.

n	$\gamma^{a}_{\hat{\beta}_0}$			$\gamma^{a}_{\hat{\beta}_1}$			*Nominal*
	True $\Delta = 0.5$ *Specified* Δ			*True* $\Delta = 0.5$ *Specified* Δ			
	0.3	0.5	0.7	0.3	0.5	0.7	
10	0.801	0.802	0.803	0.893	0.895	0.898	0.950
20	0.870	0.871	0.871	0.928	0.931	0.935	0.950
50	0.899	0.899	0.900	0.932	0.940	0.950	0.950
100	0.912	0.913	0.913	0.924	0.942	0.960	0.950
200	0.923	0.924	0.925	0.908	0.943	0.970	0.950
500	0.933	0.934	0.934	0.870	0.947	0.986	0.950
1000	0.939	0.939	0.940	0.827	0.952	0.994	0.950
2000	0.939	0.939	0.940	0.783	0.958	0.998	0.950
5000	0.944	0.944	0.944	0.744	0.960	1.000	0.950

[a] Standard errors of $\gamma_{\hat{\beta}_0}$ and $\gamma_{\hat{\beta}_1}$ are nominally 0.001.

true standard deviation ratio is $\delta = S_Y/S_X = 2.0$, (2) the estimation on the sample is done with an incomplete knowledge of δ, modelled as a uniform distribution over the interval between 1.0 and 3.0 ($\Delta = 0.5$), and (3) the bootstrap CI construction is allowed to mis-specify the incomplete knowledge by letting δ^* be uniformly distributed over the intervals between 1.4 and 2.6 (specified $\Delta = 0.3$) or between 0.6 and 3.4 (specified $\Delta = 0.7$). Specifying $\Delta = 0.3$ (instead of the correct $\Delta = 0.5$) constitutes a case of overestimation of the accuracy of the prior knowledge, $\Delta = 0.7$ means an underestimation and $\Delta = 0.5$ is an unbiased estimation.

The first result is that such a mis-specification has no effect on the accuracy of CIs for the intercept. Table 8.7 displays results (for $\Delta = 0.3$, 0.5 and 0.7) that are, within the bounds of the "simulation noise," indistinguishable.

The second result is that mis-specified prior knowledge has a clear effect on the accuracy of CIs for the slope. Table 8.7 shows results for $\Delta = 0.3$ and 0.7 to deviate from those for the correct value of $\Delta = 0.5$. If we underestimate the accuracy of the prior knowledge about the size of the measurement standard deviations ($\Delta = 0.3$ instead of 0.5), then the

CIs become too narrow and the coverage is reduced; if we overestimate the accuracy ($\Delta = 0.7$ instead of 0.5), then the CIs become too wide and the coverage is inflated.

8.4 Example: climate sensitivity

The effective climate sensitivity, denoted here as Λ_S^{-1}, is a parameter that relates changes in annual-mean surface temperature to changes in the radiative forcing (greenhouse gases, etc.) of the climate system. Its units are °C (or K) per Wm^{-2}. Climate sensitivity may vary with forcing history and climatic state, reflecting the influence of varying feedback mechanisms (amplifying or attenuating) in the climate system (Mitchell et al. 1987). The lack of an accurate knowledge of Λ_S^{-1} in the recent past (since, say, 1850) is one of the major obstacles for making accurate projections of future temperatures by means of AOGCMs (Forster et al. 2007).

The traditional estimation method for Λ_S^{-1} seems to be via perturbed climate models experiments, where the temperature response of the system is studied for a range of variations of model parameters and forcing scenarios (Forster et al. 2007). Due to the limited performance of climate models, it may be helpful to consider estimations that are based entirely on direct observations. We therefore relate variable $Y(i)$, the observed temperature changes from 1850 to 2001, to variable $X(i)$, the radiative forcing variations. The time series with standard errors are shown in Fig. 8.6. Since the predictor (forcing) has been determined with error, our model is the linear errors-in-variables regression (Eq. 8.1). The estimation objective is the slope, $\beta_1 = \Lambda_S^{-1}$.

The result (Fig. 8.7) from WLSXY estimation of Λ_S^{-1} is $0.85\ \mathrm{K\,W^{-1}m^2}$. The 95% CI, a Student's t interval obtained from pairwise-MBBres resampling with $B = 2000$ (Algorithm 8.1), is [$0.47\ \mathrm{K\,W^{-1}m^2}$; $1.24\ \mathrm{K\,W^{-1}m^2}$]. The 90% CI, a level often used in the IPCC–WG I Report's chapter on radiative forcing (Forster et al. 2007), is [$0.53\ \mathrm{K\,W^{-1}m^2}$; $1.17\ \mathrm{K\,W^{-1}m^2}$]. The climate literature often uses the "equilibrium climate sensitivity," which is defined as the temperature change that would be approached in a (hypothetical) equilibrium following a doubling of the atmospheric "equivalent carbon dioxide concentration" (representing all greenhouse gases). This other sensitivity value is around $(4\ \mathrm{W^{-1}m^2})\Lambda_S^{-1}$, at least in the climate world of the E-R AOGCM of the National Aeronautics and Space Administration Goddard Institute for Space Studies, New York (Foster et al. 2008). Thus, the WLSXY result suggests that a CO_2 doubling will lead to a temperature increase of 3.4 K.

What are the effects of autocorrelation? The block bootstrap resampling took into account the relatively strong memory of temperature and

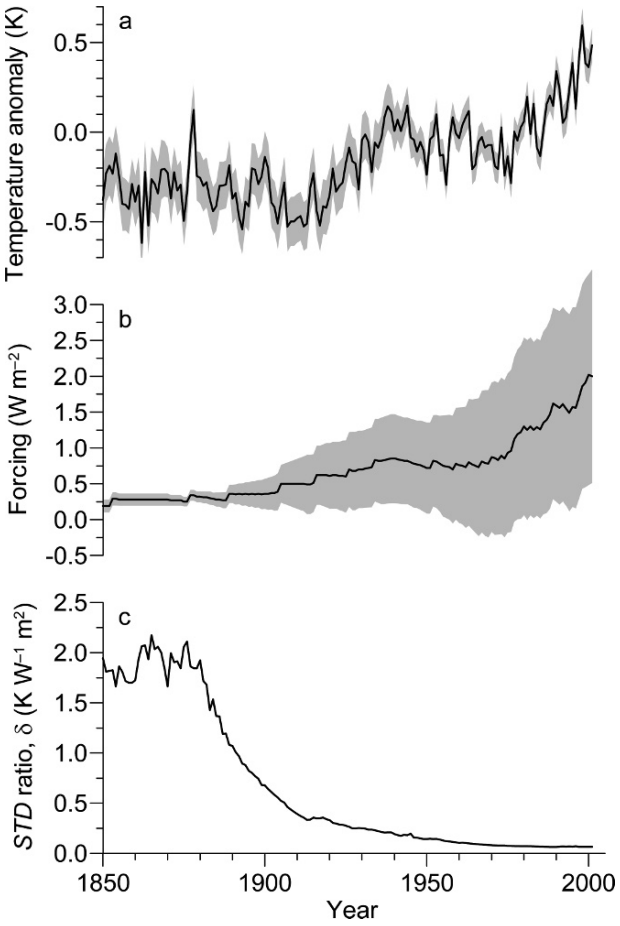

Figure 8.6. Northern hemisphere temperature anomalies and climate forcing, 1850–2001: data. **a** The temperature time series, $y(i)$, is shown (*solid line*) as deviation from the 1961–1990 average ($n = 152$). The annual-mean composite was derived using instrumental data from several thousand stations on land and sea (HadCRUT3 data set). The temperature standard error, $s_Y(i)$, is shown (*shaded band*) as $\pm 2 s_Y(i)$ interval around $y(i)$; it reflects following sources of uncertainty (Brohan et al. 2006): measurements, reporting, inhomogeneity correction, sampling, station coverage and bias correction of sea-surface temperatures. **b** The radiative forcing time series, $x(i)$, is shown (*solid line*) with $\pm 2 s_X(i)$ uncertainty band (*shaded*); it comprises following components thought to influence temperature changes (Hegerl et al. 2006; Forster et al. 2007): changes of atmospheric concentrations of greenhouse gases, solar activity variations (Fig. 2.12) and changes of sulfate and other aerosol constituents in the troposphere (lower part of the atmosphere). **c** Standard deviation ratio, $\delta = s_Y(i)/s_X(i)$. (Data from (**a**) Brohan et al. (2006) and (**b**) Hegerl et al. (2006).)

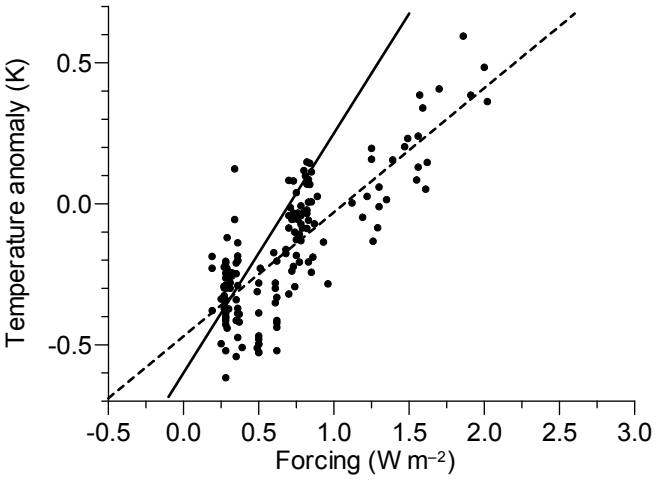

Figure 8.7. Northern hemisphere temperature anomalies and climate forcing, 1850–2001: fit. WLSXY estimation yields a straight regression line (*solid*) with a slope (i.e., effective climate sensitivity) of $\widehat{\beta}_1 = 0.85\ \mathrm{K\,W^{-1}m^2}$. Also shown is OLS regression line (*dashed*).

forcing noise components ($\widehat{a}'_X = \widehat{a}'_Y = 0.82$) by selecting a block length of $l = 18$. Ignoring autocorrelation (setting $l = 1$) would make the CI too narrow; for example, the 90% CI would become $[0.56\ \mathrm{K\,W^{-1}m^2}; 1.14\ \mathrm{K\,W^{-1}m^2}]$.

The estimate and, more, the CI for Λ_S^{-1} should be assessed, however, with caution.

- CI construction (Algorithm 8.1) used pairwise-MBBres resampling with an assumed constant standard deviation ratio of $\delta = 0.66$ (time-average). This was done because of the absence of Monte Carlo tests of adaptions of pairwise-MBBres resampling with respect to heteroscedastic errors. Instead we imposed an uncertainty of δ measured by the "incomplete prior knowledge" parameter Δ (Eq. 8.11). The employed value of $\Delta = 0.5$ may have been too small and produced a too narrow CI. Particularly, unrecognized temperature variations not caused by measurement error or forcing changes, that is, "internal temperature variability," may let δ increase and reduce the sensitivity estimate (Laepple T 2010, personal communication). Note that WLSXY estimation itself recognized heteroscedasticity.
- The HadCRUT3 temperature data (Brohan et al. 2006) are down-biased between about 1940 and the mid-1960s because of an unrecognized change in 1945 in the sea-surface measurement techniques

(Thompson et al. 2008). Since this interval is short relative to the total observation interval, the influence of the inhomogeneity on the Λ_S^{-1} estimate should be small.

- The tropospheric aerosol component of the forcing is known only with a "low" to "medium–low" scientific understanding (Forster et al. 2007). The aerosol contribution to $X(i)$ and $S_X(i)$ may be large in error. Consequently, the error in $S_X(i)$ and δ may be large, and the parameter Δ may be larger than 0.5 (or even another model of the incomplete prior knowledge required). We stress that the large error of the predictor necessitates fitting an errors-in-variables regression model. Ignoring this error (i.e., using OLS estimation) would strongly underestimate the climate sensitivity (Fig. 8.7).
- Volcanic eruptions, providing large negative forcing components (cooling) have been ignored in the estimation (because of the many unknowns), although the observed temperature time series (Fig. 8.6a) includes this effect. Since the number of large eruptions during the 151-year interval (Hegerl et al. 2006) is assessed as relatively small (about 8 eruptions with < -2.0 Wm^{-2} in the northern hemisphere), this omission should have a minor influence on the Λ_S^{-1} estimate.
- Ocean heat uptake has similarly been ignored, although observed temperatures may show this influence. Assuming that it cannot be neglected would (1) increase the Λ_S^{-1} estimate and (2) widen its CI.
- The analysis focused on the temperature of the northern hemisphere, while the concept of climate sensitivity applies to the globe. The superiority of temperature data quality for the northern part (more stations) suggested this restriction. Obviously, other geographic parts, including the globe, may be analysed in an analogous manner.

8.5 Prediction

A prediction is a statement about an uncertain event. In climate sciences the events lie often in the future (forecast) but frequently also in the past (hindcast), see the introductory examples (p. 3). In the context of the present chapter, we wish to predict an unobserved value, $y(n+1)$, given a sample, $\{t(i), x(i), y(i)\}_{i=1}^{n}$, and a new observation, $x(n+1)$, of the predictor variable made at time $t(n+1)$.

A typical situation is when a relation between a climate variable, $Y(i)$, and a proxy variable, $X(i)$, is to be established. Suppose we observed $\{t(i), x(i), y(i)\}_{i=1}^{n}$ over a time interval $[t(1); t(n)]$ but have available a longer proxy time series, $\{t(i), x(i)\}_{i=1}^{m}$ with $m > n$ (often $m \gg n$). If $t(i)$ denotes age and $t(i) > t(n)$ for $i > n$, then we wish "to hindcast"

$y(i)$ for $i > n$. An example is δ^{18}O as precipitation proxy; $y(i)$ is precipitation, $x(i)$ is δ^{18}O from a speleothem, $[t(1); t(n)] = [0\,\text{a};\ 50\,\text{a}]$ is the interval for which we have instrumental measurements of $y(i)$ (the past 50 years) and $[t(1); t(m)] = [0\,\text{a};\ 500\,\text{a}]$ is the interval covered by the speleothem samples (the past 500 years). If $t(i)$ denotes time, then we wish to forecast. An example is climate model projections; $y(i)$ is precipitation, $x(i)$ is modelled precipitation (AOGCM), $[t(1); t(n)] = [1950;\ 2010]$ is the interval for which we have instrumental measurements of $y(i)$ and $[t(1); t(m)] = [1950;\ 2100]$ is the interval analysed by means of the climate model (a typical value for the upper bound used by IPCC–WG I (Houghton et al. 2001; Solomon et al. 2007) in its reports).

Prediction can be performed by fitting a regression model and utilizing the estimated regression parameters. In the linear case (Fuller 1987: Section 1.6.3 therein):

$$\widehat{y}(n+1) = \widehat{\beta}_0 + \widehat{\beta}_1\, x(n+1), \tag{8.13}$$

where $\widehat{\beta}_0$ and $\widehat{\beta}_1$ have been estimated using the sample $\{t(i), x(i), y(i)\}_{i=1}^{n}$ and the new observation is $x(n+1)$.

Which method is suitable for estimating $\widehat{\beta}_0$ and $\widehat{\beta}_1$?

Fuller (1987: pp. 75–76 therein) explains that usage of OLS, ignoring measurement errors of the predictor, is justified when $x(n+1)$ is drawn from the same distribution that generated $\{x(i)\}_{i=1}^{n}$. This means effectively that two conditions have to be met:

1. $S_X(n+1) \cdot X_{\text{noise}}(n+1)$ has the same properties (range, shape, etc.) as $S_X(i) \cdot X_{\text{noise}}(i), i = 1, \ldots, n$;
2. $X_{\text{true}}(n+1)$ has the same properties as $X_{\text{true}}(i), i = 1, \ldots, n$.

Fuller advises further to take measurement error into account when not both conditions are satisfied. This can be done, for example, by using WLSXY (or OLSBC) estimation. Treating the regression estimates as if they were known parameters, Fuller (1987: p. 76 therein) gives the following expression for the prediction standard error:

$$\widehat{\text{se}}_{\widehat{Y}(n+1)} = \left[S_Y(n+1)^2 + \widehat{\beta}_1^2\, S_X(n+1)^2 \right]^{1/2}. \tag{8.14}$$

We argue that in climatology the above conditions are almost exclusively not satisfied, and we advise to use WLSXY (or OLSBC) as a more conservative approach. In the majority of applications, $t(n+1)$, the time value related to the new measurement, is outside of $[t(1); t(n)]$, and $x(n+1)$ does not necessarily originate from a random drawing from the process $X_{\text{true}}(i), i = 1, \ldots, n$. The new measurement may rather

constitute a step in a new direction of the course of climate, and it is safer to allow for that possibility by using WLSXY (or OLSBC).

However, the "machine error bar" (Eq. 8.14) may be too small because it does not include the estimation errors of the regression parameters. Therefore, it is advisable to use the bootstrap prediction error:

$$\widehat{\mathrm{se}}_{\widehat{Y}(n+1)} = \left\{ \sum_{b=1}^{B} \left[\widehat{Y}^{*b}(n+1) - \left\langle \widehat{Y}^{*b}(n+1) \right\rangle \right]^2 \Big/ (B-1) \right\}^{1/2}, \quad (8.15)$$

where $\left\langle \widehat{Y}^{*b}(n+1) \right\rangle = \sum_{b=1}^{B} \widehat{Y}^{*b}(n+1)/B$ and the determination of $\widehat{Y}^{*b}(n+1)$ is explained (sample level) within Algorithm 8.1.

Another source of prediction error, difficult to quantify, stems from the extrapolation. This regards (1) the standard deviations, $S_X(n+1)$ and $S_Y(n+1)$, under heteroscedasticity but also (2) the possibility that with $x(n+1)$ outside of the observation interval, from $\min(x(i))$ to $\max(x(i))$, or with $t(n+1)$ outside of $[t(1); t(n)]$, new laws set in and, if unrecognized, may bias the prediction. A physical theory behind the regression model may guard against such errors (background material).

8.5.1 Example: calibration of a proxy variable

Calibrating a proxy variable, $X(i)$, means quantifying the relation with a climate variable, $Y(i)$, by means of regression. Since $X(i)$ is usually observed with measurement error, the errors-in-variables equation (8.1) has to be considered. The fitted regression curve serves for predicting an uncertain value, $y(n+1)$, given a new proxy measurement, $x(n+1)$. Calibration is ubiquitous in quantitative paleoclimatology. Examples: oxygen isotopic composition in a marine sediment core is a proxy for temperature (Fig. 1.2), hydrogen isotopes in an ice core indicate temperature (Fig. 1.3a). Here we look at δ^{18}O in a coral as a proxy for air temperature.

We make two further remarks. First, the calibration methodology applies also to predicting future climate values by means of climate models. Second, the core of the interest lies usually in relative variations, changes of a variable—the slope (which itself is susceptible to estimation bias).

Draschba et al. (2000) calibrated δ^{18}O, measured in a coral taken from a site off the coast of Bermuda, against observations of air-temperature on that island (Fig. 8.8). The calibration curve, established for the time interval from 1856 to 1920, was then used to make a hindcast of temperature for the interval from 1350 to 1630 (by using measurements from another coral located close to the site of the "calibration coral").

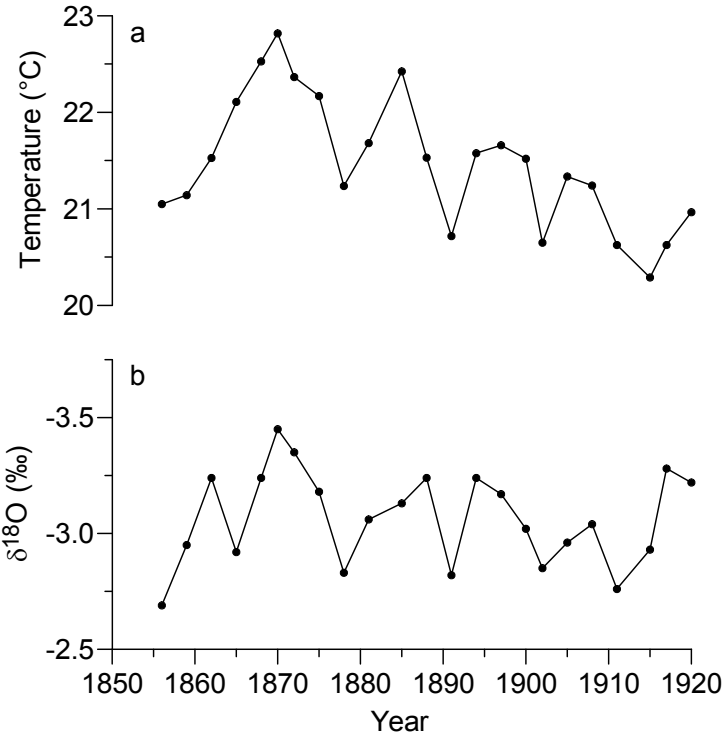

Figure 8.8. Bermuda air temperature and coral δ^{18}O, 1856–1920: data. **a** The annual-mean temperature time series, $y(i)$, is shown only at those time points for which δ^{18}O values (**b**) are available ($n = 23$). The temperature standard error, $s_Y(i)$, is assumed to be constant and equal to 0.03°C (Table 1.3). **b** The δ^{18}O time series, $x(i)$, is unevenly spaced ($d(i) = 2$ or 3 a). The δ^{18}O values have a constant measurement error of $s_X = 0.07‰$ (Draschba et al. 2000). (The temperature data are digitized values from Draschba et al. (2000: Fig. 2c therein), the δ^{18}O data were downloaded from `http://doi.pangaea.de/10.1594/PANGAEA.88200` (17 September 2009).)

The accuracy of the δ^{18}O timescale, crucial for a successful calibration, is excellent owing to the presence of seasonal density banding (visible on X-ray photographs). Measurement procedures and errors (Fig. 8.8) are described in detail by Draschba et al. (2000). Sample material requirements led to an unevenly spaced δ^{18}O time series, with $D'(i) = D(i) = d(i) = 2$ or 3 a (see Fig. 1.13 for definitions). Draschba et al. (2000) transformed the temperature record (monthly observations) by binning to either an annual resolution or a 3-year resolution. Their

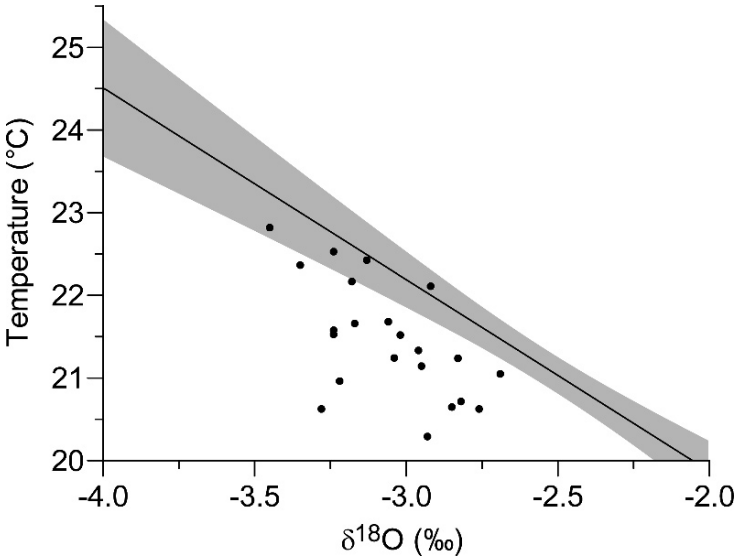

Figure 8.9. Bermuda air temperature and coral δ^{18}O, 1856–1920: prediction. OLSBC estimation yields a straight prediction line (*solid*) with an intercept of $\widehat{\beta}_0 = 15.2°\text{C}$ and a slope of $\widehat{\beta}_1 = -2.3°\text{C‰}^{-1}$. Also shown is 95% Student's t confidence band (*shaded*), obtained from bootstrap resampling (pairwise-MBBres with $\widehat{\tau}'_X = \widehat{\tau}'_Y = 6.9\,\text{a}$, $l = 6, B = 2000$).

calibration result did not strongly depend on that choice. Here we use the annual values from those years for which also δ^{18}O values exist (Fig. 8.8).

The calibration curve (Fig. 8.9) has a slope that is in size larger by a factor of approximately 1.3 than that estimated by Draschba et al. (2000). This considerable deviation is likely the result of an ignored bias correction in the original paper. The bias-corrected OLSBC fit curve deviates considerably from a na-ive per-eye fit through the points (Fig. 8.9). (Interestingly, the authors considered already their slope estimate as rather large in absolute size.) The pointwise bootstrap confidence band allows to quantify the prediction uncertainty, also outside of the original range of observations (Fig. 8.9).

The bootstrap prediction error (Eq. 8.15), averaged over the interval of x values shown in Fig. 8.9, is equal to 0.25°C, while the "machine error bar" (Eq. 8.14) is 0.16°C.

Two further remarks ought to be made. First, the confidence band assumes a time-independent calibration relation and homoscedastic er-

rors. This assumption may be violated. Second, the confidence band may be inaccurate owing to the limited data size, as the Monte Carlo experiments (Section 8.3) show.

8.6 Lagged regression

Let us reconsider the linear errors-in-variables model (Eq. 8.1) in continuous time, T. Assume for convenience homoscedasticity. Introduce a time lag parameter, H, to shift the predictor variable, such that

$$Y(T) = \beta_0 + \beta_1 \left[X(T+H) - S_X \cdot X_{\text{noise}}(T+H)\right] + S_Y \cdot Y_{\text{noise}}(T). \quad (8.16)$$

A lag $H > 0$ ($H < 0$) means that variations of "true" Y lead over (lag behind) variations of "true" X. This is a lagged errors-in-variables regression model.

Measured time series are discrete in time and finite in size. Assume for convenience even time spacing ($d(i) = d = \text{const.}$) and introduce a dimensionless time lag, $h = H/d$, such that

$$Y(i) = \beta_0 + \beta_1 \left[X(i+h) - S_X \cdot X_{\text{noise}}(i+h)\right] + S_Y \cdot Y_{\text{noise}}(i), \quad (8.17)$$

$i = 1, \ldots, n-h$. Given a bivariate sample, $\{x(i), y(i)\}_{i=1}^{n}$, the task is to estimate β_0, β_1 and h.

WLSXY estimation should in principle be possible by minimizing a normalized sum,

$$SSQWXY(\beta_0, \beta_1, h) = (n-h)^{-1} \sum_{i=1}^{n-h} \frac{[y(i) - \beta_0 - \beta_1 x(i+h)]^2}{S_Y^2 + \beta_1^2 S_X^2}. \quad (8.18)$$

This may be achieved technically by numerical minimization (Section 8.8) of $SSQWXY(\beta_0, \beta_1, \widetilde{h})$ for a fixed (candidate) lag, $\widetilde{h}$, and a brute-force search over a range of $\widetilde{h}$ values. Intuitively, if $1 \ll h \ll n$, then the error due to the discretization of the time should be smaller than when h is close to either bound.

A more general, realistic situation arises when the two time series were observed at mutually unequal times. This has been explored in the context of correlation estimation (Section 7.5), where the time gaps could be bridged owing to the presence of persistence. The situation becomes even more realistic (difficult), when timescale errors are introduced. We analyse such an example (Section 8.6.1), where we resort to interpolation. The taken approach is somewhat ad-hoc. The theoretical knowledge about estimators and their properties, let alone CI construction, is rather limited for such situations, and the given literature (background material) does not cover this issue exhaustively.

8.6.1 Example: CO_2 and temperature variations in the Pleistocene

One of the major contributions of ice cores as climate archives is information about CO_2 variations far back in time (late Pleistocene). The Vostok core's record, first drilled and measured over the past 160 ka (Barnola et al. 1987), was later extended to the full span of 420 ka (Petit et al. 1999). The longest CO_2 record currently available (past 800 ka) comes from the EPICA Dome C ice core (Siegenthaler et al. 2005; Lüthi et al. 2008). The major finding from those ice core studies was that not only temperature and ice volume underwent large changes during the ice age (100-ka cycle), but also the atmospheric CO_2 concentration. We explore here the full Vostok span of changes of CO_2 and temperature (inferred via δD), shown in Fig. 1.3, to estimate the phase relations between these changes. Such relations constitute a basis for erecting a causal climatological theory of the late Pleistocene ice age—which does not yet exist in sufficient detail. We follow the paper by Mudelsee (2001b), who used lagged regression as a tool for phase relationship estimation.

Mudelsee (2001b) deviated in some technical points from the errors-in-variables methodology developed in the previous sections. These and some additional points are discussed first, the results shown thereafter.

First, the time values of the predictor variable, $\{t_X(i)\}_{i=1}^{3311}$, are not identical to those of the response variable, $\{t_Y(j)\}_{j=1}^{283}$. Allowing for a candidate lag, $\widetilde{H}$, requires a time shift. For those reasons, the lag estimation used linear interpolation of the x values (Fig. 8.10), $t_X = t_Y(j) - \widetilde{H}$. The fact that the lag is imposed for computational reasons on $t_Y(j)$ rather than $t_X(i)$ (Eq. 8.16), is not relevant for the estimation.

Second, the lagged regression employed a parabolic model. This performed slightly better than the linear one, as evaluated by means of the reduced sum of squares (fourth point). (A logarithmic model would yield similar values as the parabolic (Fig. 8.12).)

Third, the predictor's error has an upper limit of $s_X = 1‰$ (Petit et al. 1999), which is clearly smaller than the standard deviation (spread) of the $x(i)$ values of 17‰. This means only a small estimation bias when ignoring measurement error (Mudelsee 2001b).

Fourth, the estimation (Mudelsee 2001b) used GLS (Section 4.1.2) with the $\mathbf{V}$ matrix elements given by $\exp[-|t_Y(j_1) - t_Y(j_2)|/\tau_Y]$. Because the persistence time, τ_Y, was unknown, a second brute-force loop for τ_Y was nested and the overall minimum taken as solution. The resulting reduced least-squares sum is

$$SSQG_\nu(\boldsymbol{\beta}, H, \tau_Y) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \, \mathbf{V}^{-1} \, (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \, / \nu, \tag{8.19}$$

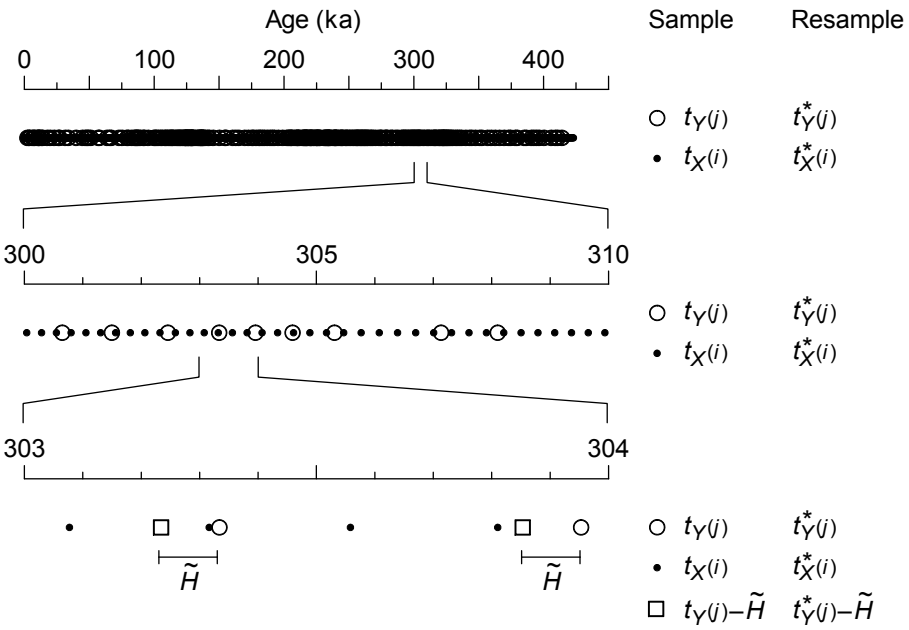

Figure 8.10. Vostok deuterium and CO_2, timescales for lag estimation. The time interval [303 ka; 304 ka] illustrates the relation between the predictor (X) variable, deuterium, and the response (Y) variable, lagged CO_2. The candidate lag in time is $\widetilde{H}$. The predictor values are obtained by linear interpolation of the x and x^* values, $t_X = t_Y(j) - \widetilde{H}$ and $t^*_X = t^*_Y(j) - \widetilde{H}$, respectively. (Original data shown in Fig. 1.3.)

where

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \text{ (parameter vector)}, \tag{8.20}$$

$$\mathbf{y} = \begin{bmatrix} y(1) \\ \vdots \\ y(n-h) \end{bmatrix} \text{ (response vector)}, \tag{8.21}$$

$$\mathbf{X} = \begin{bmatrix} 1 & x'(1) & x'(1)^2 \\ \vdots & \vdots & \vdots \\ 1 & x'(n-h) & x'(n-h)^2 \end{bmatrix} \text{ (predictor matrix)}, \tag{8.22}$$

$\nu = n - h - 3$ (degrees of freedom) and x' is interpolated x (Fig. 8.10). The linear model has no β_2 parameter and $\nu = n - h - 2$. The step size of the brute-force search for $\widehat{H}$ was 5 a.

The resulting lag estimate is $\widehat{H} = -1.3$ ka, that is, a lag of CO_2 variations behind temperature variations. The resulting persistence time is $\widehat{\tau}_Y = 0.92$ ka. The reduced least-squares sum in dependence on $\widetilde{H}$ is shown in Fig. 8.11. The resulting parabolic fit is shown in Fig. 8.12.

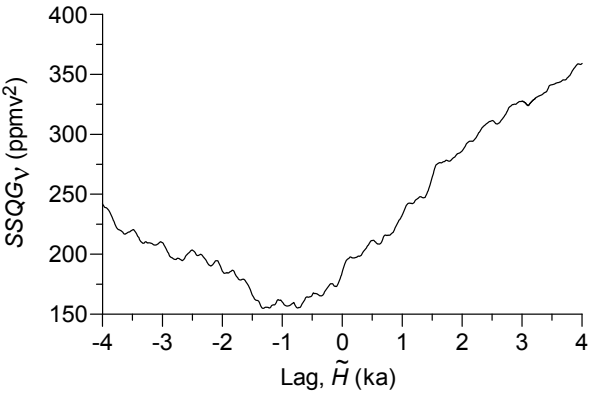

Figure 8.11. Vostok deuterium and CO_2, reduced sum of squares. The minimum (i.e., lag estimate) is at $\widetilde{H} = -1.3$ ka. (After Mudelsee 2001b.)

Both predictor and response (x, y) exhibit measurement and proxy errors, and both timescales (t_X, t_Y) show dating uncertainties. These four error sources propagate into the estimation standard error of the lag, $\widehat{\mathrm{se}}_{\widehat{H}}$. Mudelsee (2001b) determined $\widehat{\mathrm{se}}_{\widehat{H}}$ by means of a parametric surrogate data approach (Algorithm 8.2).

The first error source (x) was simulated (Mudelsee 2001b) as

$$x^*(i) = x(i) + x_{\mathrm{noise}}(i), \tag{8.23}$$

where $x_{\mathrm{noise}}(i)$ is a realization of a Gaussian AR(1) process with standard deviation $s_X = 1.0‰$ (Petit et al. 1999) and persistence time $\tau_X = 2.1$ ka (Chapter 2).

The second error source (y) was simulated analogously as

$$y^*(i) = y(i) + y_{\mathrm{noise}}(i), \tag{8.24}$$

where the noise process had a standard deviation $s_Y = 2.5$ ppmv (Petit et al. 1999) and persistence time $\tau_Y = \widehat{\tau}_Y = 0.92$ ka.

The third error source (t_X) was simulated (Mudelsee 2001b) using the depth points of the ice core samples (Petit et al. 1999) and a non-parametric fit of the "sedimentation rate" (Fig. 4.13). The simulated

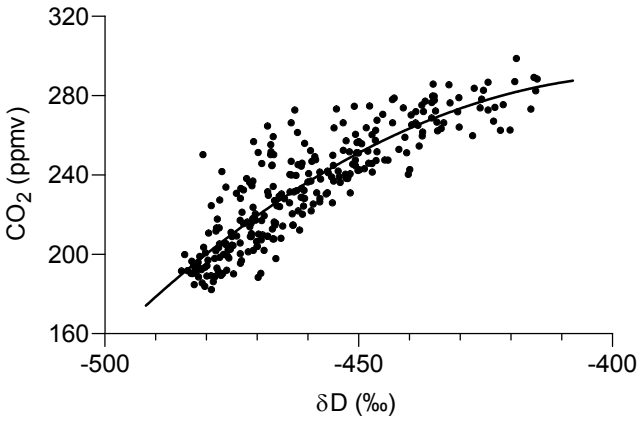

Figure 8.12. Vostok deuterium and CO_2, parabolic fit. Data points are lagged CO_2 ($\widehat{H} = -1.3$ ka) against interpolated δD ($n - h = 280$). The fit line is given by $y = -1482 - 9.05x - 0.012x^2$. (After Mudelsee 2001b.)

Step 1	Time series	$\{t_X(i), x(i)\}_{i=1}^{n_X}$,
		$\{t_Y(j), y(j)\}_{j=1}^{n_Y}$
Step 2	Lag estimate	$\widehat{H}$
	via minimization of $SSQG_\nu(\boldsymbol{\beta}, H, \tau_Y)$	
Step 3	Simulated time series;	$\{t_X^{*b}(i), x^{*b}(i)\}_{i=1}^{n_X}$,
	b, counter	$\{t_Y^{*b}(j), y^{*b}(j)\}_{j=1}^{n_Y}$
Step 4	Replication	$\widehat{H}^{*b}$
Step 5	Go to Step 3 until $b = B$ (usually $B = 2000$)	
	replications exist	$\{\widehat{H}^{*b}\}_{b=1}^{B}$
Step 6	Calculate standard error and CIs	

Algorithm 8.2. Determination of bootstrap standard error and construction of CIs for lag estimate in lagged regression, surrogate data approach (Sections 3.3.3 and 3.4). The algorithm is applicable also to other estimation techniques than $SSQG_\nu$ minimization (Step 2).

sedimentation rate was obtained parametrically (Mudelsee 2001b) by imposing a relative, Gaussian error of 1.2%. The simulated sedimen-

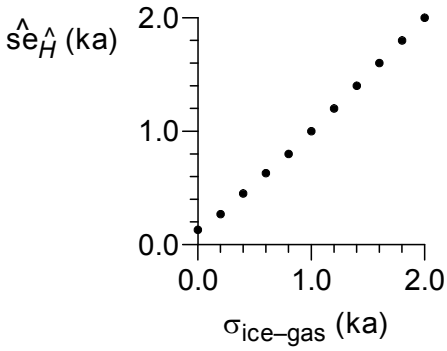

Figure 8.13. Vostok deuterium and CO_2, sensitivity study of lag estimation error.

tation rate, combined with the depth points, resulted in a simulated timescale (Section 4.1.7). In a final step, the simulated timescale was randomly compressed or expanded to fit into the GT4 timescale error range (Petit et al. 1999), which is ≤ 5 ka for the last 110 ka, ≤ 10 ka for "most of the record" (interpreted as 110–300 ka by Mudelsee (2001b)) and ≤ 15 ka in the early part.

The fourth error source (t_Y) was simulated on basis of the simulated ice-ages (t_X^*). The additional error contribution comes from the uncertainty in the ice–gas age difference,

$$t_Y^* = t_X^* + \mathcal{E}_{\mathrm{N}(0,\,\sigma^2_{\mathrm{ice-gas}})}(\cdot). \tag{8.25}$$

Petit et al. (1999: p. 434 therein) reported $\sigma_{\mathrm{ice-gas}}$ to be 1 ka or more.

The surrogate data approach yielded (Mudelsee 2001b) $\widehat{\mathrm{se}}_{\widehat{H}} = 1.0$ ka. To restate, the lag estimation result is that temperature variations occurred 1.3 ± 1.0 ka before CO_2 variations.

The crucial point for achieving such a small estimation error is that x and y were measured on the same core (Vostok). This means a rather close coupling of t_X^* and t_Y^* (Eq. 8.25). Only the uncertainty in the ice–gas age difference weakens the coupling. Had CO_2 been measured on a core from a different site, no coupling would exist and t_X^* and t_Y^* had to be simulated independently of each other, leading to a clearly larger lag estimation error than in the present case.

The lag estimation result underlines the importance of the uncertainty, $\sigma_{\mathrm{ice-gas}}$, in the ice–gas age difference, which contributes nearly 100% to the lag estimation error of 1.0 ka. A sensitivity study (Fig. 8.13) quantifies this contribution over a range of prescribed $\sigma_{\mathrm{ice-gas}}$ values. For example, in the case of $\sigma_{\mathrm{ice-gas}} = 0.2$ ka the lag estimation error would be $\widehat{\mathrm{se}}_{\widehat{H}} = 0.27$ ka. In the case of a perfectly known difference

($\sigma_{\text{ice–gas}}$ equal to zero), the remaining error sources would propagate into $\widehat{\text{se}}_{\widehat{H}} = 0.13$ ka.

As regards causal explanations of the late Pleistocene glacial cycles, Mudelsee (2001b) noted that Vostok's air temperature (δD) represents, at best, the southern hemisphere and that there exists a time lag of the variations relative to the northern hemisphere (Blunier et al. 1998). However, the complexity of the ice-age climate may be better understood, that is, the set of feasible causal scenarios (Broecker and Henderson 1998: Table 1 therein) further constrained, with the help of quantified phase relations.

8.7 Background material

OLSBC estimation of the slope has also been denoted as attenuation-corrected OLS (ACOLS) estimation (Ammann et al. 2009).

The **method of moments** estimator of the standard deviation of the predictor in the case of homoscedasticity, S_X, is (Fuller 1987: Eq. (1.3.10) therein):

$$\widehat{S}_X = (2\delta)^{-1} \left\{ m_{YY} + \delta m_{XX} - \left[(m_{YY} - \delta m_{XX})^2 + 4\delta m_{XY}^2 \right]^{1/2} \right\}, \tag{8.26}$$

where

$$\delta = \lambda^{1/2} = S_Y / S_X, \tag{8.27}$$

the moments are

$$m_{YY} = \sum_{i=1}^{n} [y(i) - \bar{y}]^2 / (n-1) , \tag{8.28}$$

$$m_{XX} = \sum_{i=1}^{n} [x(i) - \bar{x}]^2 / (n-1) , \tag{8.29}$$

$$m_{XY} = \sum_{i=1}^{n} [x(i) - \bar{x}] [y(i) - \bar{y}] / (n-1) \tag{8.30}$$

and the sample means are

$$\bar{y} = \sum_{i=1}^{n} y(i) / n \tag{8.31}$$

and

$$\bar{x} = \sum_{i=1}^{n} x(i)\,/n\ . \tag{8.32}$$

$\widehat{S}_X$ is plugged in for S_X (Eq. 8.6).

WLSXY estimation of a linear relationship between two variables that are both subject to error has been studied, and the geometric interpretation been made, already before and at the beginning of the twentieth century (Adcock 1877, 1878; Pearson 1901); see also Wald (1940) and Fuller (1987, 1999). The method to fit a hyperplane to data with errors in all their coordinates (possibly more than two) is also denoted as total least squares (Nievergelt 1998).

Non-Gaussian, heteroscedastic noise components in the linear errors-in-variables regression model can be taken into account in the estimation using GLS, that is, using the covariance matrix, analogously to Section 4.1.2. In practical applications to climatological problems, where the covariance matrix is unknown and has to be estimated, an iterative procedure may be used. Fuller (1987: Section 3.1 therein) describes GLS estimation for serially independent noise components and gives a result (standard errors of parameters) that is valid for large data sizes. He advises to consider developing a model for the error structure if the data size is small. However, it is not clear whether such a classical approach to parameter error determination can be applied also to serially dependent noise components.

Correlated noise components in the linear errors-in-variables regression model can be taken into account. York (1969) adapts a least-squares criterion to recognize correlation between $X_{\text{noise}}(i)$ and $Y_{\text{noise}}(i)$ and gives an example from radiometric dating. Freedman (1984) and Freedman and Peters (1984) present two-stage regression with bootstrap resampling as a method to treat a correlation between $X_{\text{noise}}(i)$ and $Y(i)$. Fuller (1987: Section 3.4 therein) presents a transformation for dealing with correlation between $X_{\text{noise}}(i)$ and $X(i)$.

Multiplicative measurement error may occur in form of $X(i) = X_{\text{true}}(i) \cdot X'_{\text{noise}}(i)$, where the primed noise component is dimensionless. Carroll et al. (2006: Section 4.5 therein) mention transformation methods that may be applied in this case.

Nonlinear errors-in-variables models can be estimated on basis of several assumptions about the model and the noise properties, by using numerical techniques for solving the maximum likelihood or least-squares optimizations (Fuller 1987: Section 3.3 therein). A recent book (Carroll et al. 2006) gives more details.

The **pairwise-MBBres** algorithm from Section 8.2 is a response to resolving the "quite nonstandard" (Hall and Ma 2007: p. 2621 therein) situation, where neither the true predictor variable, $X_{\text{true}}(i)$, nor the errors, $X_{\text{noise}}(i)$, "can be directly accessed." Previously, Efron and Tibshirani (1993: Section 9.5 therein) and Davison and Hinkley (1997: Section 6.2.4 therein) considered that pairwise bootstrap resampling is applicable to errors-in-variables regression problems. Linder and Babu (1994) presented another alternative to the simple pairwise resampling. These authors scaled the residuals in both dimensions (X, Y) and resampled independently from both sets. They analysed maximum likelihood estimation with known standard deviation ratio and tested the accuracy of bootstrap CIs (percentile and Student's t) by means of Monte Carlo experiments, finding acceptable levels of accuracy. This was confirmed in an additional simulation study of slope estimation (Musekiwa 2005) with small data sizes ($n = 20, 30$). It should be interesting to investigate further the approach of Linder and Babu (1994), adapted to the climatologically more realistic situation where the standard deviation ratio is not exactly known and the errors exhibit serial dependence.

The **approaching of finite RMSE values** or, equivalently, the absence of shrinking CIs with $n \to \infty$ was verified for slope estimation and falsified for intercept estimation (Section 8.3.3). Previously, Booth and Hall (1993) found a non-shrinking bootstrap confidence band in a Monte Carlo experiment on prediction (Section 8.5), $\widehat{y}(n+1) = \widehat{\beta}_0 + \widehat{\beta}_1\, x(n+1)$. Thus, it appears that this observation (Booth and Hall 1993) has its origin in the non-shrinking of $\text{RMSE}_{\widehat{\beta}_1}$.

The **simulation–extrapolation** algorithm (Carroll et al. 2006: Chapter 5 therein) is a bias correction method based on Monte Carlo simulations. The idea is to add artificial measurement error to the data and study the dependence of an estimate (say, $\widehat{\beta}_1$) in dependence of the size of the artificial error. Extrapolation to zero size should, so the idea, provide an unbiased estimate.

Prediction and forecasting by means of regression and other models is reviewed by Ledolter (1986). The success of prediction depends, of course, on the validness of the regression model and the absence of disturbing "latent" variables (Box 1966). A physical theory behind the model is a guard against wrong conclusions based on such disturbances. For example, radiation physics and meteorology support the concept of climate sensitivity (estimated by means of a regression of changes in temperature on changes in radiative forcing, see Section 8.4) and refute claims that time acts as a latent variable. On the other hand, a re-

gression of annual temperature on the annual output of scientific papers on global warming over the past, say, 150 years, would find a strong relation (highly significant slope)—however, a spurious relation because the latent variable time is acting and no theory exists that supports the model.

Lagged regression as presented in Section 8.6 (that means, with one single lag parameter, h), is a special case of rational distributed lag models (Koyck 1954; Dhrymes 1981; Doran 1983; Pankratz 1991), where

$$Y(i) = \beta_0 + \beta_{1,0}X(i) + \beta_{1,1}X(i+1) + \beta_{1,2}X(i+2) + \cdots + S_Y \cdot Y_{\text{noise}}(i). \tag{8.33}$$

The sequence $\{\beta_{1,0}, \beta_{1,1}, \beta_{1,2}, \ldots\}$ is called impulse response function. Note that the equation ignores errors in the predictor. Fitting such models may be performed using maximum likelihood (Dhrymes 1981) or frequency-domain techniques (Hannan and Robinson 1973; Hannan and Thomson 1974; Hamon and Hannan 1974; Foutz 1980). This technique is frequently applied in econometrics. One of the rare applications to climatology is the work by Tol and de Vos (1993), who estimated a lagged regression of annual-mean temperature, 1951–1979, on atmospheric CO_2 concentration. Insofar climate is a dispersive system, where the response of one variable on the input of another is frequency-dependent, it should be worth developing further such models and fitting techniques that take into account typical properties of paleoclimatic series (measurement errors, unequal times and uncertain timescales).

The **effective climate sensitivity** is usually denoted as λ_S^{-1}. Various estimation approaches have been carried out, Table 8.8 gives a short overview. The approach via the heat capacity (Schwartz 2007) opened an interesting exchange of arguments in the Journal of Geophysical Research. Let C denote the effective heat capacity (change in heat per change in temperature) per unit area that is coupled to the relevant timescale of a perturbation (i.e., years to decades). The perturbation regards the radiative balance of the Earth (change in forcing). Schwartz (2007) estimated C (with standard error) to be 17 ± 7 W a m^{-2}K^{-1}. The C value reflects mainly the upper part of the ocean, which can take up heat on the discussed timescale of (anthropogenically enhanced) radiative perturbations. The simple equation,

$$\tau = C \cdot \Lambda_S^{-1}, \tag{8.34}$$

describes the time span (relaxation, τ) over which a radiative perturbation (C) has an effect on the temperature (Λ_S^{-1}). Schwartz (2007) estimated τ by fitting an AR(1) model (Chapter 2) to observational data. The criticism on this approach (Foster et al. 2008; Knutti et al.

2008; Scafetta 2008) was centred on the AR(1) model as over-simplistic and estimation bias. In his reply, Schwartz (2008) kept the AR(1) model but conceded τ to be larger (8.5 ± 2.5 a) than in his original contribution (5 ± 1 a). The revised estimate for τ leads to the entry in Table 8.8.

Table 8.8. Estimates of the effective climate sensitivity.

Λ_S^{-1} *Estimate*[a] ($K\,W^{-1}m^2$)	*Approach*	*Reference*
0.29 $[0.05; 0.53]^{b,c}$	Direct observations, 2000–2006	Chylek et al. (2007)
0.48 $[0.24; 0.72]^{b,d}$	Direct observations, 2000–2006	Chylek et al. (2007)
0.51 $[-0.01; 1.03]^{b}$	Physical principles (heat capacity)	Schwartz (2008)
0.65 $[0.09; 1.21]^{b}$	Thermodynamical model	Scafetta and West (2007)
0.70 $[0.38; 1.55]^{e}$	Climate model and observations, 1000–2001	Hegerl et al. (2006)
0.85 $[0.53; 1.17]^{c}$	Direct observations, 1850–2001	This book
1.53 $[0.40; \infty]^{e,f}$	Direct observations, 1861–1900 and 1957–1994	Gregory et al. (2002)

[a] With 90% CI.
[b] CI calculated as ±2 standard error interval.
[c] Ignoring ocean heat uptake.
[d] Assuming strong ocean heat uptake.
[e] Calculated from originally estimated equilibrium climate sensitivity.
[f] Median given instead of estimate.

The **leads and lags of carbon dioxide** variations relative to those of temperature in the Pleistocene have been studied by several researchers on time series from ice cores from Antarctica. Previously to Mudelsee (2001b), who estimated $\widehat{H} = -1.3 \pm 1.0$ ka (a lag of CO_2), the original authors of the 0–420 ka Vostok paper (Petit et al. 1999) had found, seemingly by per-eye inspection, that CO_2 decreases lag behind temperature decreases by several ka. Cuffey and Vimeux (2001: p. 523 therein) believed that the lag, "especially during the onset of the last glaciation, about 120 ka ago," is largely an "artefact caused by variations of climate in the water vapour source regions." They presented model simulations that correct for this effect and lead to $\widehat{H} \approx 0$ ka using the Vostok data, 0–150 ka and 150–350 ka. Subsequently, Monnin et al. (2001) analysed high-resolution records ($\bar{d} \approx 0.18$ ka for CO_2) from the EPICA Dome C site over the interval 9–22 ka by means of an explorative technique ("correlation maximization") similar to the brute-force search (Section 8.6.1). They obtained an estimate of $\widehat{H} = -0.41$ ka, which was assessed as not significant owing to the uncertainty of the ice–gas age difference. Caillon et al. (2003) revisited the Vostok ice core, inspected

the time interval around Termination III (230–255 ka) and took argon isotopes instead of deuterium as proxy for temperature changes. The motivation for performing the new measurements was the idea that the poorer proxy quality of argon isotopes would be more than compensated by the smaller timescale uncertainties. Since argon is, as CO_2, contained in the enclosed air bubbles in the ice, no uncertainty of the ice–gas age difference influences lag estimation (Eq. 8.25). The result of correlation maximization (Caillon et al. 2003) was a lag of CO_2, $\widehat{H} = -0.8 \pm 0.2$ ka, where the error bar value is based on an assessment of the uncertainty of the ice accumulation (but not on an additional consideration of the proxy errors). The "EPICA challenge" (Wolff et al. 2005), issued to paleoclimatologists, was to predict CO_2 concentration for the interval 420–740 ka on basis of the then available EPICA Dome C deuterium (temperature) and dust records (EPICA community members 2004). The simple model, lagged regression of Vostok CO_2 on EPICA deuterium (temperature), calibrated on the 0–420 ka records, did not produce the worst of the eight predictions, as was found when the EPICA Dome C CO_2 series became known. The complete interval back to 800 ka from the EPICA ice core archive of past changes in temperature (Jouzel et al. 2007) and CO_2 (Siegenthaler et al. 2005; Lüthi et al. 2008) enables to analyse also temporal changes of the lag, H, concepts that the ice-age climate relationships changed for a while after a glacial termination (Raynaud et al. 1993). To summarize, the overall lag of CO_2 variations behind temperature variations during the late Pleistocene appears significant. This is a quantitative basis for developing and testing concepts about causes and effects of long-term climate changes (Broecker and Henderson 1998; Saltzman 2002), about how the external astronomical forcing (Milankovitch cycles) propagates into the geographic regions, and how the climate variables respond. Further refining the ice-age theory is currently an active research field (Kawamura et al. 2007; Huybers and Denton 2008; Wolff et al. 2009). It is important to note that the characteristic timescales on which the analysed Pleistocene climate changes occurred, are relatively long: the average spacing ($\bar{d}$), the estimated lag ($\widehat{H}$) and its estimation error ($\widehat{\mathrm{se}}_{\widehat{H}}$) are between several hundred and a few thousand years. The late Pleistocene lag estimates are therefore hardly relevant as regards concepts about the ongoing climate change, which is anthropogenically enhanced since, say, 150 years. This recent change is considerably faster than the late Pleistocene change, it leads to CO_2 levels not experienced during at least the past 800 ka and it affects other physical processes. The consideration from physics and meteorology that the recent change has a positive time lag (CO_2 rise

before temperature rise) is not contradicted by the finding that the late Pleistocene change had a negative time lag. The scientifically interesting question is whether humans are able to put a (temporary) end to the succession of glacials and interglacials (Berger and Loutre 2002).

Errors-in-variables regression models for climatology have not often been formulated in an explicit manner in the research literature. Allen and Stott (2003) showed theoretically the importance of an unbiased slope estimation for linear models that relate temperature changes predicted by an AOGCM with observed temperature changes. Hegerl et al. (2007a) studied in that manner proxy calibration to reconstruct 30–90°N mean annual land-surface temperature for the past 1500 years. Kwon et al. (2002) fitted a nonlinear model to dating samples,

$$Y(i) = \frac{\exp\left(\lambda_{^{40}\mathrm{K}} \cdot \tau_{\mathrm{FCs}}\right) - 1}{\exp\left\{\lambda_{^{40}\mathrm{K}}\left[X(i) - S_X \cdot X_{\mathrm{noise}}(i)\right]\right\} - 1} + S_Y \cdot Y_{\mathrm{noise}}(i), \quad (8.35)$$

$i = 1, \ldots, n$. They used five paired samples of $Y(i) = {}^{40}\mathrm{Ar}/{}^{39}\mathrm{Ar}$ ratio and $X(i) =$ reference age, observed with small, homoscedastic (S_X, S_Y), mutually independent measurement errors of assumed Gaussian shape. The estimation parameters were $\lambda_{^{40}\mathrm{K}}$ (decay constant of $^{40}\mathrm{K}$) and τ_{FCs} (age of Fish Canyon sanidine age standard). Kwon et al. (2002) derived maximum likelihood estimators and analysed bootstrap standard errors based on the surrogate data approach. Their Monte Carlo study showed that the estimates do not strongly depend on the Gaussian assumption. The result, $\widehat{\lambda}_{^{40}\mathrm{K}} = 5.4755 \pm 0.0170 \cdot 10^{-10}\ \mathrm{a}^{-1}$, leads to a half-life estimate (Section 1.6) of $\widehat{T}_{1/2} = \ln(2)/\widehat{\lambda}_{^{40}\mathrm{K}} = 1.266 \pm 0.004$ Ga. Bloomfield et al. (1996) made a multivariate nonlinear regression of daily tropospheric ozone concentration in the Chicago metropolitan area, 1981–1991, on a variety of predictors, including temperature, wind speed and relative humidity. The interesting point in the context of this chapter is that also lagged predictors (H prescribed as 1 or 2 days) were included. Bloomfield et al. (1996) used GLS estimation (Gallant 1987: Sections 2.1 and 2.2 therein) and obtained parameter standard errors by means of jackknife resampling (Section 3.8), which takes serial dependence into account.

8.8 Technical issues

WLSXY minimization of $SSQWXY(\beta_0, \beta_1)$ is numerically difficult because the slope, β_1, appears in the denominator of the least-squares sum (Eq. 8.8). The routine Fitexy (Press et al. 1992) parameterizes the slope as $\beta_1' = \tan^{-1}(\beta_1)$, scales the y values and uses Brent's search (Section 4.5) with a starting value for the slope from an initial OLS estimation. (A more recent Numerical Recipes edition is Press et al.

(2007).) Papers on the way from the work of Deming (1943) and York (1966) to the routine Fitexy include Reed (1989, 1992) and Squire (1990). This book follows those authors in usage of WLSXY for estimation, but it does not so for parameter error determination; for that purpose it uses instead bootstrap resampling. Extensions of WLSXY to nonlinear regression problems (nonlinear in the parameters) were considered by Jefferys (1980, 1981) and Lybanon (1984). A review of least-squares fitting when both variables are subject to error (Macdonald and Thompson 1992) studied besides WLSXY also other weighting techniques. It appears that a generalized least-squares estimation method for the case of (1) serial correlations in both X and Y errors and (2) correlation between X and Y errors, supported by a proof of optimality (in terms of, say, RMSE) under the Gaussian distributional assumption, has not yet been developed.

LEIV1 is another Fortran implementation of WLSXY estimation (York 1966), available at `http://lib.stat.edu/multi/leiv1` (26 October 2009).

LEIV3 is a Fortran software for maximum likelihood fitting of linear errors-in-variables models with heteroscedastic noise components (Ripley and Thompson 1987), available at `http://lib.stat.edu/multi/leiv3` (26 October 2009).

Bootstrap software for errors-in-variables regression is not abundant. Carroll et al. (2006) and Hardin et al. (2003) mention Stata software, available from the site `http://www.stat.tamu.edu/~carroll` (26 October 2009). Software for block bootstrap resampling seems to be unavailable—limiting the ability to study errors-in-variables regression with autocorrelated noise components.